## SDG 4 DATA DIGEST 2018

# Data to Nurture Learning

United Nations
Educational, Scientific and
Cultural Organization

UNESCO
INSTITUTE
FOR
STATISTICS

SUSTAINABLE
DEVELOPMENT
GOALS

SDG 4 DATA DIGEST

# Data to Nurture Learning

United Nations
Educational, Scientific and
Cultural Organization

UNESCO INSTITUTE FOR STATISTICS

SUSTAINABLE DEVELOPMENT GOALS

**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 Organization currently has 195 Member States and 11 Associate Members.

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 2018 by:

UNESCO Institute for Statistics
P.O. Box 6128, Succursale Centre-Ville
Montreal, Quebec H3C 3J7 Canada

Tel: +1 514-343-6880
Email: uis.publications@unesco.org
http://www.uis.unesco.org

ISBN 978-92-9189-230-3
Ref: UIS/2018/ED/SD/9



Photo credits: Arne Hoel, Aigul Eshtaeva and Visual News Associates/World Bank

# Table of contents

## List of boxes



## List of figures

## List of tables

# Foreword

Education is one of a nation's greatest assets and the foundation for strong and peaceful societies. However, illiteracy and low educational achievement are persistent challenges for many developing countries, for international agencies, for global educational programmes and for the achievement of the world's education goals.

The Millennium Development Goals (MDGs) adopted by world leaders in 2000 created greater awareness of the state of education in developing countries and the massive efforts needed to achieve the MDG targets of universal access to primary education, as well as full global literacy and numeracy.While major strides were made on access to education by the 2015 deadline for the MDGs, the quality of education remained a major concern.

In 2015, the Sustainable Development Goals (SDGs) set out new ambitions for education, with SDG 4 requiring a quality education from pre-primary to upper secondary level of education for every child by 2030. The global commitment to improving education captured in SDG 4 aims to address an educational crisis, with more than 617 million children and adolescents unable to read a simple sentence or handle a basic math calculation.

Today, we are faced with three major issues: there are many children who are still out of school and who have little chance of acquiring basic skills in reading and mathematics; there are children who are enrolled in school but at risk of leaving before they gain these skills; and the continuing and pervasive problem of poor quality education. This is why SDG 4 includes targets to ensure improvements in the quality of teaching, the inclusion of skills for a modern and increasingly digital society and ensuring that children and youth are not only in the classroom, but also learning.

As the custodian agency for SDG 4 indicators, the UNESCO Institute for Statistics (UIS) is leading the development of the methodologies and standards needed to produce internationally-comparable indicators. Based on this foundation, the UIS is working with national statistical offices, line ministries and international organizations worldwide to track global progress on education while creating the frameworks and tools for effective monitoring at national, regional and global levels.

The 2018 edition of the *SDG 4 Data Digest: Data to Nurture Learning*, builds on last year's report, which proposed a conceptual framework and tools to help countries improve the quality of their data and fulfil their reporting requirements. In this report, we present the wide range of national and cross-national learning assessments currently underway and the assessment experiences of practitioners in the field. The report draws on these experiences to present pragmatic approaches that can help countries monitor progress and make the best possible use of data for policymaking purposes.

As this report shows, we do not need to create entirely-new monitoring mechanisms: we can build on what is already in place. For example, we are making great strides towards reporting on Indicator 4.1.1 on the proportion of children and young people at three different stages of their education who have a minimum proficiency level in reading and mathematics, thanks to existing national, regional and cross-national assessments.

Through the Global Alliance to Monitor Learning (GAML), we are working with countries, assessment agencies, donors and civil society groups to take a harmonised approach to data collection, setting benchmarks and enhancing quality control to ensure the effective use of results to improve learning. This

is both a technical and political process that will take time and money to perfect.

As shown in the Digest, data on learning outcomes are a necessity, not a luxury, needed by every country. On average, low- and middle-income countries require about US$60 million per year to regularly assess learning. These costs are really investments that will yield exponential benefits for the current generation and those to come.

**Silvia Montoya**
**Director**
**UNESCO Institute for Statistics**

# Acknowledgements

This report would not have been possible without the remarkable contributions of experts representing a wide range of institutions.

We are grateful to the following experts for submitting invaluable background papers and other contributions for this report:

**Chapter 2 and Annexes**

| | |
|---|---|
| | IEA |
| Dana Kelly | Technical Director (MSI) |
| Jeff Davis | Technical Director (MSI) |
| Abdullah Ferdous | Technical Director (MSI) |

**Chapter 3**

| | |
|---|---|
| Hilaire Hounkpodoté | PASEC Coordinator (CONFEMEN) |
| Paulína Koršňáková | Senior Research and Liaison Adviser (IEA) |
| Dirk Hastedt | Executive Director (IEA) |
| Michael Ward | Senior Analyst, Development Co-operation Directorate (OECD) |
| Élisé Wendlassida Miningou | Education Economist (GPE Secretariat) |
| Ramya Vivekanandan | Senior Education Specialist (GPE Secretariat) |
| Luis Crouch | Senior Economist, International Development Group (RTI International) |
| Amber Gove | Director, Research (RTI International) |
| Rukmini Banerji | Director (ASER Centre) |
| Suman Bhattacharjea | Director of Research (ASER Centre) |
| James Ciera | Senior Data Analyst (Twaweza East Africa) |
| Sara Ruto | Director (PAL Network) |
| Mary Goretti Nakabugo | Twaweza Country Lead and Regional Manager (Uwezo East Africa) |

**Chapter 4**

| | |
|---|---|
| Claudia Cappa | Statistics and Monitoring Specialist (UNICEF) |
| Nicole Petrowski | Statistics and Monitoring Officer (UNICEF) |
| Magdalena Janus | Professor of Psychiatry and Behavioural Neurosciences, Offord Centre for Child Studies (McMaster University) |

**Chapter 5**

| | |
|---|---|
| Manos Antoninis | Director (GEMR) |
| Yves Punie | Senior Scientist (European Commission Joint Research Centre) |
| Riina Vuorikari | Researcher (European Commission Joint Research Centre) |
| Marcelino Cabrera | Researcher (European Commission Joint Research Centre) |
| Nancy Law | Professor, Centre for Information Technology (CITE), University of Hong Kong |
| Mart Laanpere | Senior Researcher (Tallinn University) |

**Chapter 6**

| | |
|---|---|
| William Thorn | Senior Analyst (OECD) |
| Koji Miyamoto | Senior Economist (World Bank) |
| Madina Bolly | Coordinator (RAMAA) |

**Chapter 8**

| | |
|---|---|
| Paulína Koršňáková | Senior Research and Liaison Adviser (IEA) |
| Dirk Hastedt | Executive Director (IEA) |
| Elena Arias Ortiz | Education Senior Associate (IDB) |
| Florencia Jaureguiberry | Education Consultant (IDB) |
| Pablo Zoido | Education Lead Specialist (IDB) |

We would also like to thank Luis Crouch for his careful review and dedication to this report, as well as Maria José Ramírez, who helped to coordinate the contributions from external authors and provided them with feedback.

The report was edited by Barbara Zatlokal. The report was coordinated and supported by the Learning Outcome unit of the UIS, which is being led by Silvia Montoya, UIS Director and acting head of section, with the assistance of Omneya Fahmy and Adolfo Imhof. Brenda Tay-Lim of the UIS provided valuable collaboration in Chapter 6. Katja Frostell of the UIS communications unit coordinated the production of the report.

# Acronyms and abbreviations

| | |
|---|---|
| A4L | Assessment for Learning |
| ACER | Australian Council for Educational Research |
| ADEA-NALA | Association for the Development of Education in Africa – Network of African Learning Assessments |
| ALL | Adult Literacy and Life Skills Survey |
| ANCEFA | Africa Network Campaign on Education for All |
| ASER | Annual Status of Education Report |
| BFI | Big Five Inventory |
| CEPAL | United Nations Economic Commission for Latin America and the Caribbean |
| CIMA | *Centro de Información para la Mejora de los Aprendizajes* |
| CITE | Centre for Information Technology in Education (Hong Kong University) |
| CLA | Citizen-led assessment |
| CNA | Cross-national assessment |
| CONFEMEN | *Conférence des ministers de l'Education des Etats et gouvernements de la Francophonie* (Conference of Ministers of Education of States and Governments of Francophonie) |
| DART | Data Alignment Record Tool |
| DCPs | Developing Country Partners |
| DESI | Digital Economy and Society Index (European Union) |
| DHS | Demographic and Health Survey |
| DIA | Development in the Americas |
| DigComp | Digital Competence Framework for Citizens |
| DigCompEdu | Digital Competence Framework for Educators |
| DigCompOrg | Digitally Competent Educational Organizations |
| DLGF | Digital Literacy Global Framework |
| EAP-CDS | East Asia-Pacific Child Development Scales |
| ECD | Early childhood development |
| ECDI | Early Childhood Development Index |
| EDI | Early development instrument |
| EFA | Education for All |
| EGMA | Early Grade Mathematics Assessment |
| EGRA | Early Grade Reading Assessment |
| EHCI | Early Human Capability Index |
| ELA | Early Learning Assessment |
| ELDS | Early Learning Development Standards |
| ePIRLS | Progress in International Reading Literacy Study online |
| EQAP | Educational Quality and Assessment Programme |
| EQF | European Qualification Framework |
| ESCS | Economic, social and cultural status |
| ESD | Education for sustainable development |
| ESPIG | Education sector programme implementation grants |
| ETS | Educational Testing Service |
| EU | European Union |
| FCAC | Fragile or conflict-affected countries |
| GAML | Global Alliance to Monitor Learning |

| | |
|---|---|
| GCED | Global Citizenship Education |
| GEM | Global Education Monitoring |
| GPE | Global Partnership for Education |
| GRA | Global and regional activities |
| HCF | Harmonised Competency Framework |
| IAEG-SDG | Inter-Agency and Expert Group on SDG Indicators |
| IALS | International Adult Literacy Survey |
| IBE | International Bureau of Education |
| IC3 | Certiport Internet and Computing Core Certification |
| ICCS | International Civic and Citizenship Study |
| ICDL | International Computer Driving License |
| ICFES | Colombian Institute for Educational Evaluation |
| ICILS | International Computer and Information Literacy Study |
| ICT | Information and communications technology |
| ICU | International Communications Union |
| IDB | Inter-American Development Bank |
| IDELA | International Development and Early Learning Assessment |
| IEA | International Association for the Evaluation of Educational Achievement |
| IRT | Item response theory |
| ISCED | International Standard Classification of Education |
| ITU | International Telecommunication Union |
| IVQ | Survey on Information Exchange and Daily Life |
| JRC | Joint Research Centre (European Commission) |
| KGPE | Knowledge and Good Practice Exchange |
| KIX | Knowledge and Innovation Exchange |
| LAMP | Literacy Assessment and Monitoring Programme |
| LANA | Literacy and Numeracy Assessment |
| LLECE | Latin American Laboratory for Assessment of the Quality of Education |
| MDGs | Millennium Development Goals |
| MELQO | Measuring Early Learning Quality and Outcomes |
| MICS | Multiple Indicator Cluster Surveys |
| MIRT | Multidimensional Item Response Theory |
| MODEL | Measurement of Development and Early Learning |
| MSA | Modern Standard Arabic |
| MSI | Management Systems International |
| NA | National assessment |
| NAEP | National assessment of educational progress |
| NAS | National Achievement Survey |
| NEQMAP | Network on Education Quality Monitoring in Asia Pacific |
| NESPAP | National education systems and policies in Asia-Pacific |
| NGO | non-governmental organization |
| NLA | National learning assessment |
| NSAT | National standardised achievement test |
| OECD | Organisation for Economic Co-operation and Development |
| OPCE | Plurinational Observatory of Educational Quality |
| PASEC | *Programme d'analyse des systems* éducatifs *de la CONFEMEN* (Programme of Analysis of Education Systems of CONFEMEN) |
| PDPs | Policy definitions of performance |

| | |
|---|---|
| PIAAC | Programme for the International Assessment of Adult Competencies (OECD) |
| PILNA | Pacific Islands Literacy and Numeracy Assessment |
| PIRLS | Progress in International Reading Literacy Study |
| PISA | Programme for International Student Assessment |
| PISA-D | Programme for International Student Assessment for Development |
| PLDs | Performance level descriptors |
| PPP | Purchasing power parity |
| PRIDI | Regional Project on Child Development Indicators |
| PRIMR | Primary Mathematics and Reading Programme |
| R&D | Research and development |
| RAMAA | Action Research on Measuring Literacy Programme Participants' Learning Outcomes |
| REDUCA | *Red Latinoamericana por la Educatión* |
| REESAO | *Réseau pour l'excellence de l'enseignement supérieur en Afrique de l'Ouest* |
| SABER | Systems Approach for Better Education Results |
| SACMEQ | Southern and Eastern Africa Consortium for Monitoring Educational Quality |
| SDG | Sustainable Development Goal |
| SEA-PLM | Southeast Asia Primary Learning Metrics |
| SELFIE | Self-reflection on Effective Learning by Fostering Innovation through Educational Technologies |
| SPC | Pacific Community |
| STEP | Skills towards Employability and Productivity |
| TAG | Technical Advisory Group |
| TALENT | Teaching and Learning Educators' Network for Transformation |
| TaRL | Teaching at the Right Level |
| TCG | Technical Cooperation Group on the Indicators for SDG 4-Education 2030 |
| TERCE | *Tercer Estudio Regional Comparative y Explicativo* (Third Regional Comparative and Explanatory Study) |
| TIMSS | Trends in International Mathematics and Science Study |
| UC | University of Chile |
| UIL | UNESCO Institute for Lifelong Learning |
| UIS | UNESCO Institute for Statistics |
| UNICEF | United Nations Children's Fund |
| USAID | United States Agency for International Development |
| WHO | World Health Organization |

# Introduction

According to new estimates from the UNESCO Institute for Statistics (UIS), more than 617 million children and adolescents are not be able to read or handle mathematics proficiently. About two-thirds of these children and youth are in school, some of them dropping out before reaching the last grade of the cycle (UIS, 2017g). This highlights the critical need to improve the quality of education while expanding access to ensure that no one is left behind.

Not only is the learning crisis alarming from a national, social and economic perspective, but it also threatens the ability of individuals to climb out of poverty through better income-earning opportunities. Greater skills not also raise their potential income, but well-educated individuals are also more likely to make better decisions – such as vaccinating their children – and educated mothers are more likely to send their own children to school. The learning crisis is, simply, a massive waste of talent and human potential. For this reason, many of the global goals depend on the achievement of Sustainable Development Goal 4 (SDG 4), which demands an inclusive and equitable quality education and the promotion of "lifelong learning opportunities for all".

UIS data suggest that the numbers are rooted in three common problems. First, a lack of access, with children who are out of school having little or no chance of reaching a minimum level of proficiency; second, failure to keep every child on track and proceeding through the system on time and retaining them in school; and third, the issue of the quality of education and what is happening within the classroom itself.

## FROM "OUT OF SCHOOL" TO "CHILDREN NOT LEARNING" AND "SKILLS SHORTAGE"

The number of out-of-school children (or its effective complement, the net enrolment rate) became, in many respects, the de facto flagship indicator during the Education for All (EFA) and Millennium Development Goals (MDGs) era. The most visible change in the Education 2030 and SDG era is the more explicit focus on the quality of education. In practice, for monitoring purposes, this is increasingly interpreted through learning outcomes.

All the evidence suggests that we are far from meeting the targets stated in SDG 4. In sub-Saharan Africa, for example, out-of-school children represent a relatively high proportion (46%) of the total number of children not achieving the minimum proficiency in reading. The proportion for adolescents is 65%. While this particular example shows that close to one-half of children not learning are out of school, this is not the case for other regions. Western Asia and Northern Africa, as well as Central Southern Asia, have around 20% of children not learning as out-of-school children. This number is quite alarming, since it indicates that 80% of children not able to achieve minimum proficiency levels are in classrooms but not learning. If the majority of children and adolescents not learning are actually in school, this means that policies need to address improving the quality of the education offered.

Estimates show that two-thirds (68%) of these children – or 262 million out of 387 million – are in school and will reach the last grade of primary education but will not achieve minimum proficiency levels in reading. These findings show the extent to which education systems around the world are failing to provide a quality education and decent classroom conditions in which children can learn.

Another 78 million (20%) are in school but are not expected to reach the last grade of primary education. Unfortunately, according to UIS data, 60% of the dropout occurs in the first three grades of the school cycle, leaving many children without foundational skills. While there are many reasons for high dropout rates, the data underscore the need to improve education policies by tailoring programmes to meet the needs of different types of students, especially those living in poverty. The benefits of education must outweigh the opportunity costs of attending school for students and their households.

It is not surprising to find that 40 million children (10% of the total) unable to read proficiently have either left school and will not re-enrol or have never been in school and will probably never start. If current trends continue, they will remain permanently excluded from the basic human right of education.

Finally, there are roughly another 21 million children of primary school age who are currently not in school but are expected to start late. About 6.9 million of these children will not reach the last grade of primary education and are therefore not expected to achieve minimum proficiency levels in reading.

While the numbers are staggering, they show the way forward. Two-thirds of the children and youth not learning are actually in school. We can reach these children. But not by simply hoping that they stay in school and grasp the basics. We must understand their needs and address the shortcomings of the education currently on offer.

## LEARNING AND SDG 4 ON EDUCATION

Learning is paramount for all the sustainable development goals. It is needed to end poverty, ensure prosperous and fulfilling lives in harmony with nature, and to foster peaceful, just and inclusive societies. Learning is a process that happens throughout the whole life cycle, from when we are born until we die. We learn to walk, to talk, to think, to love and to care for others. We learn the social values that allow us to live together. We learn the working skills needed to make a living and to contribute to society. We learn to learn.

However, sustainable development is at risk when a vast proportion of the world's population is not learning: for instance, when infants and young children do not learn to play with each other via skills such as impulse control, when children do not learn to read and think mathematically or critically, and when young people and adults do not learn the digital skills needed to function in modern societies.

Because learning is so critical for our lives and the future of our planet, a global commitment has been made to monitor and support learning. SDG 4 on education is at the core of this effort. Many indicators are directly related to learning:

- Indicator 4.1.1: Proportion of children and young people: (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex.
- Indicator 4.2.1: Proportion of children under 5 years of age who are developmentally on track in health, learning and psychosocial well-being, by sex.
- Indicator 4.4.2: Percentage of youth and adults who have achieved at least a minimum level of proficiency in digital literacy skills.
- Indicator 4.6.1: Percentage of the population in a given age group achieving at least a fixed level of proficiency in functional (a) literacy and (b) numeracy skills, by sex.
- Indicators 4.7.1, 4.7.4 and 4.7.5: Percentage of students by age group (or education level) showing adequate understanding of issues relating to global citizenship and sustainability and percentage of 15-year-old students showing proficiency in knowledge of environmental science and geoscience.

Other SDG 4 or related indicators, such as the completion of an education cycle or the transition to

the next cycle, are strongly affected by learning levels. This is the case for completion rates and out-of-school rates just to mention a few.

## BUILDING ON CURRENT PRACTICES

The UN's adoption of indicators focusing on the attainment of specific proficiency levels through education raises exciting and complex questions on how the UIS, as the custodian agency with the mandate to complete the methodological development of most of the SDG 4 indicators, will move forward in the measurement and reporting of learning. The approach promoted by the Institute will have far-reaching implications not just for the quality and relevance of international statistics but also for how more than 200 national education authorities measure learning and improve access to quality education, while supporting teaching and learning in the classroom.

Political commitments and investments have already been made according to preferences and priorities. This obviously influences what choices can be made in the future. Substantive work has been done in the learning domains that are relevant to SDG 4, but many challenges are still ahead. Work has been done in conceptualising the learning domains to be measured in the context of SDG 4, the tools to measure learning and the administration of these tools in different countries. Mathematics and reading measurement seems to be considerably far ahead, whereas other learning domains are at an earlier stage of development to inform SDG 4. There are promising initiatives to measure early child development, digital skills and work skills in the adult population. However, their coverage is more limited largely because the number of countries that regularly collect (or report) information on these domains is much lower.

Much has already been written about optimal approaches and the factors that should influence the choices. It is now abundantly clear that determining a global data collection strategy is a technically complex matter, with serious cost and behavioural implications at various levels and with many solidly entrenched points of view. Furthermore, both the political agendas and monitoring frameworks of the SDGs and Education 2030 are extremely ambitious. They demand an unprecedented increase in the collection, processing and dissemination of data from and, most importantly, within countries.

One important argument is that **the comparability of national statistics over time should receive more attention**. Until now, much of the focus has fallen on the comparability of proficiency statistics **across assessment programmes and countries.** The latter is important especially for reasons of equity between and within countries. The focus on the comparability of national statistics over time is vital in terms of UNESCO's commitment to global progress and implies somewhat different strategies to those associated with improving comparability across countries. One can think of good comparability in statistics over time, combined with a relatively crude degree of cross-country comparability, as a second-best option which can still guide global strategies in powerful ways.

Producing statistics which are comparable over programmes and countries is perhaps even more difficult than is often assumed. One reason for this is that different parts of the world have different traditions when it comes to the complexity of the items used to measure whether students are meeting various proficiency benchmarks at different grades. Some countries apply more stringent items to measure certain proficiency levels than others. The reverse seems to be the case in the upper primary grades (Gustafsson, 2018). This obviously makes it more difficult to reach a global consensus around proficiency benchmarks. Reality further complicates comparisons across countries as comparison in some points mean different years of formal schooling than in others (for instance, some countries finish primary education in the fourth grade, while others finish in the sixth grade or at the end of lower secondary school).[1]

1 Based on an analysis of International Standard Classification of Education (ISCED) levels, which provide a comprehensive framework for organizing education programmes and qualifications by applying uniform and internationally-agreed definitions to facilitate comparisons of education systems across countries (UIS, 2016b).

It is important to recognise that several years will be required to resolve all the methodological and political issues needed to report on SDG indicators on the same scale. The challenges are primarily due to the fact that learning assessment initiatives use different definitions of performance levels and, importantly, different levels of difficulty in the items that test whether proficiency is met. While discussions continue, an interim reporting strategy that maximises the use of available data has been put in place by the UIS and is discussed in this report.

## SETTING BENCHMARKS TO TRACK PROGRESS

The Education 2030 Framework for Action commits all countries to establish benchmarks for measuring progress towards SDG 4 targets using certain scales. By describing the progression of learning skills, the scales will help countries identify and agree on the benchmarks needed to define minimum proficiency levels for reporting purposes. The Global Alliance to Monitoring Learning (GAML) and the Technical Cooperation Group on the Indicators for SDG 4-Education 2030 (TCG) are leading this consensus-building process on the indicators.

The discussions on benchmarks touch every major education issue. What are the minimum levels of learning we expect children to achieve? Should there be one benchmark for developing countries and another for developed countries? Or should they be defined at the country level? Perhaps most importantly, do children and their households have the right or entitlement to a minimum level of learning?

Gathering evidence on learning is one thing; using that evidence to improve learning is another. Several authors discuss how the conceptual framework and evidence in different learning domains are being used to foster learning. For instance, the European Commission's Digital Competence Framework for Citizens (DigComp) is used to measure these skills and provide guidelines for action in education and training. Cross-national assessments have had an impact on curriculum reforms, teacher training and pedagogical resources in participating countries. National assessments are used to drive classroom reforms. In addition, a coherent international framework works best when it meshes well with coherent national frameworks, and information provided by the latter can work all the way down to the classroom level and inform formative assessment (though obviously the assessment methods are different).

Informing SDG 4 learning indicators is a necessary but not sufficient step to monitor and support learning for all. Data on learning need to be disseminated to stakeholders, both in the countries (e.g. policymakers) and in the international community (e.g. donors and international cooperation agencies). Efforts are needed to ensure that stakeholders understand, value and effectively use the information to ensure inclusive and equitable quality education for all, and that virtuous cycles of measurement/action/re-measurement are used to improve children's lives, much as has been efficiently done in other sectors.

## COSTS AND BENEFITS OF INVESTING IN LEARNING DATA

The UIS and its partners are not just interested in collecting statistics for their own sake but in establishing a data collection system and refining existing systems in a manner whereby: a) the very process of collecting data has positive side effects; and b) the statistics are used constructively to bring about better education policies that advance the SDGs.

Assessments/skills surveys required to report against SDG 4 indicators are relatively costly with respect to other data collection systems required for these indicators. It is estimated that data on the quality of learning, or proficiency levels, will account for around one-half of all costs related to SDG reporting in education (UIS, 2017e).

For instance, to report on Indicator 4.1.1, participation in one round of a large international assessment

programme (such as TIMSS[2] and PISA[3]) costs a country around US$800,000 (UIS, 2018a). The figure is lower– US$200,000 to US$500,000 – for regional cross-national programmes, such as LLECE[4] and PASEC.[5]

Given that the costs of a sample-based assessment, as well as the optimal sample size, are largely independent of the size of the country, the ratio of assessment costs to overall spending becomes higher in smaller countries. However, relative to the overall cost of providing schooling, assessment systems appear not to be costly. One can expect costs in initial cycles to be higher than in subsequent cycles due to the need for start-up and development activities.

Investment is more likely to take place if the benefits are clearly communicated. In other words, a stronger emphasis is needed on the demand for and utilisation of data, not simply supplying data (UIS, 2018a). This requires thinking differently and more broadly about processes around data. For this, human capacity is needed, both with respect to broad strategic thinking around data and also with respect to very specific skills. There is also a need for better technical documentation to guide countries. The challenge is to find the most cost-efficient, fit-for-purpose way of producing learning statistics.

## UNDERSTANDING CAPACITY DEVELOPMENT NEEDS AT THE COUNTRY LEVEL

It is worth noting that human capacity appears *under*-emphasised in the current literature on education data. In particular, human capacity to bring about innovation *within individual countries* seems under-emphasised. Instead, much of the focus falls on tools in the form of manuals and standards. These tools are important but do not guarantee, on their own, that the necessary human capacity will be built.

Cross-national assessment programmes have created networks that have facilitated country-specific capacity building. Yet the processes within these programmes are largely premised on a model where innovation and advanced technical work, for instance with respect to sampling and psychometrics, occurs in one place, while each country follows a set of instructions. The problem with insufficient innovation (as opposed to imitation) in individual countries is that country-focused *use* of the data which emerges from the cross-national programme is often limited as is capacity to design *national* programmes. Moreover, weak technical capacity in a country might mean that national assessment systems are influenced by political interference, which is a real risk in an area such as assessments.

What would probably be beneficial for capacity building is an elaborated version of a list of competences, to assist in particular developing countries to identify what skills should be developed. A “good practice“ guide provides a basic list of assessment-related skills that can be considered advanced and which, it is argued, should perhaps be secured through outsourcing. To this list can be added skills relating to the dissemination of data, such as skills in developing technical documentation accompanying the data, or metadata, and skills needed to anonymise data. Any country or education authority should aim to have these competences *within* the authority or at least the country. In other words, the aim should be to reduce the need for outsourcing. Though advanced, these skills can be considered essential for sustaining and defending an effective national assessment system.

## THE 2018 SDG 4 DATA DIGEST

This year’s edition of the SDG 4 Data Digest is dedicated to the theme of learning outcomes. It showcases the most comprehensive and up-to-date

2 Trends in International Mathematics and Science Study.
3 Programme for International Student Assessment.
4 *Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación* (Latin American Laboratory for Assessment of the Quality of Education).
5 *Programme d’analyse des systèmes éducatifs de la CONFEMEN* (Analysis Programme of the CONFEMEN Education Systems).

compilation of work relevant to inform the learning indicators of SDG 4.

The digest discusses learning evidence on early child development, mathematics and reading skills in school-aged children, and digital and work-related skills in youth and adults. It highlights the conceptual frameworks and tools developed by leading authors and institutions to understand, measure, monitor and support learning for all. It also considers the implications of reporting for SDG 4.

Chapter 1 presents the framework for reporting and data harmonisation being used by the UIS and its technical partners. This chapter defines the UIS' overall structure for all learning outcomes and skills indicators, methodological development and reporting strategy. The following chapters address the frameworks and workflows for each indicator.

Chapter 2 describes the work on Indicator 4.1.1,which deals with the proficiency of students in two learning areas (reading and mathematics) and three educational levels. The political and technical challenges and solutions are addressed and ways forward are proposed.

The following chapters describe experiences with different types and levels of assessments in various domains. The UIS did not impose a rigid structure on these chapters so as to allow the writers the opportunity to focus on areas of particular interest to a region or institution. The aim was not to be encyclopaedic, or to provide a *menu fixe*, but to allow the users to sample what the authors themselves considered the most important and useful features of their approaches. This, arguably, provides a framework for optimism: a great deal of work is already being done. At the same time, it buttresses the argument that there is still a large task ahead in terms of consolidation, finding commonalities and finding ways to link.

Chapter 3 presents the main learning assessment programmes for basic education reading and mathematics: cross-national, regional, national and population-based. Some of the cases present evidence to inform different SDG indicators whenever available. The chapter provides a fresh account of the different assessment programmes. There are several promising initiatives that will soon produce data on learning.

Chapter 4 describes three proposals that have been put forward to report on Indicator 4.2.1, which is under the custodianship of UNICEF, while chapter Chapter 5 provides a somewhat formal analysis of the current work on digital literacy measurement.

Chapter 6 discusses functional literacy and numeracy. It opens by defining the main methodological issues in comparability and charting a way forward that could include synthetic estimates and the generation of new tools as global public goods

Chapter 7 highlights the importance of national efforts to monitor learning. It provides countries with guidelines on implementing assessments, as well as SDG 4 monitoring and dissemination. It highlights the need to ensure that stakeholders have access to assessment information, understand and value it.

Chapter 8 also focuses on the dissemination of and uses for learning assessment data. It showcases how two major institutions, such as the International Association for the Evaluation of Educational Achievement (IEA), are supporting countries.

# 1. Setting a strategy to measure learning outcomes

Worldwide 617 million children and youth are not learning the basics. This has alerted the international community to the importance of tackling the assessment of learning challenges. What are "the basics"? What does "fail" mean in terms of measurement? How can we assess this number when there is no internationally-agreed methodology to do so?

Data covering all children are essential if we want to improve learning for every child and if we want to guide educational reform. The data tell us who is not learning, help us to understand why and can help to channel scarce resources to where they are most needed. A lack of learning data is an impediment to educational progress, and it is in the differences in the learning outcome levels between different groups of students that educational inequality shows up most dramatically. For example, two-thirds as many children in low-income countries complete primary schooling as in high-income countries. But, even in some middle-income countries, about 60% of children are at or below minimum learning competency levels, whereas in high-income countries there are essentially no children at this level: a difference of about 0% to 60%. Moreover, we do not even have the data for the low-income countries; we can only guess that the difference between high-income countries and the low-income countries is 0% to 80%. It is in this 80% of children learning at or below minimum competency level that global vulnerability shows most clearly.

In past years, assessing learning outcomes was not the dynamic domain it is today. There is now a profusion of assessments at international, regional and national levels, research articles are flourishing and media attention is high when new results from an international survey are published. League tables stir the debate in every country and opposition to these exercises is fierce. Concerns about country ownership and sovereignty over their own education policies are emerging, as well as questions about methodology and robustness of data. What happens behind the scenes during the production of *the* score is not always easily answered and not in ways easily understood by non-experts.

Despite this call for a strong voice to inform the debate in a neutral and meaningful way, the international community has yet to come up with a methodology to harmonise assessment programmes and ensure robust cross-country comparability, expand the number of comparison points and references for countries, and provide all citizens with a universal grid to read and understand while putting into perspective the results of any assessment.

The urgency is palpable for establishing concrete steps to obtain high-quality, globally-comparable data on learning that can be used to improve national education systems. According to the UIS, currently only one-third of countries can report on Indicator 4.1.1 with data that are partially comparable with other countries that participated in the same assessment programme. The deadline is drawing near. By the end of 2018, the education community must have a solution for how to report on SDG 4.

The education sector as a whole will be strengthened and reinforced by bringing together data on and knowledge of learning outcomes and skills from around the world through SDG monitoring. In other words, a stronger emphasis is needed on the demand and use of data, not simply the collection and supply of data (UIS, 2018b). This requires thinking differently and more broadly about the processes that are

created around data. This requires human capacity in countries, both with respect to broad strategic thinking on how to choose investments around data, how to adapt and implement them, and the very specific skills required.

It is worth noting that the need for investment in human capacity at the country level appears *under-*emphasised. In particular, human capacity to bring about innovation *within individual countries* seems under-emphasised. Instead, much of the emphasis is on tools in the form of manuals and standards. These tools are important, but on their own are not a guarantee that the necessary human capacity will be built.

Section 1.1 starts by discussing the dimensions of comparability involved in SDG reporting. Section 1.2 refers to the specific demands placed on the SDG indicators, while Sections 1.3 and 1.4 provide a brief overview of the key challenges that are common to all targets and indicators. Chapter 1 concludes by offering a framework to finalise the methodological discussion for reporting.

## 1.1 WHY AND WHAT TYPE OF COMPARABLE DATA?

A key issue in discussions relating to SDG reporting is the need to produce internationally-comparable statistics on learning outcomes. This is a challenge for various reasons. Countries may wish to use just nationally-determined proficiency benchmarks which are meaningful to the country. Even if there is the political will to adopt global proficiency benchmarks, the fragmented nature of the current landscape of cross-national and national assessment systems would make the goal of internationally-comparable statistics difficult to achieve.

More internationally-comparable statistics on learning outcomes would contribute towards a better quality of schooling around the world, and we could measure change *over time* with respect to learning outcomes and the attainment of proficiency benchmarks. If this is not done, it will not be possible to establish whether progress is being made towards the relevant SDG target, and this, in turn, will make it very difficult to determine whether strategies adopted around the world are delivering the desired results.

Improving the comparability of statistics *across* countries helps to gauge progress towards the achievement of relevant and effective learning outcomes for all young people. The logic is simple. If statistics on learning outcomes can be made comparable across countries – and more specifically across assessment programmes – at a given point in time through an equating or linking methodology, then assuming that each assessment programme produces statistics which are comparable *over time*, statistics even in future years will be comparable between countries. Global aggregate statistics will also be calculated over time which will reflect the degree of progress.

Comparable statistics across countries are important, and efforts towards global comparison have been vital for improving our knowledge of learning and schooling. But this is not the only dimension that is relevant for SDG reporting. Statistics must be comparable *across countries* (across space), and the focus must be more on the comparability of national statistics *over time.* It should be acknowledged that the two aspects, space and time, are interrelated but also to some degree independent of each other. In fact, just as good comparability of national statistics over time can co-exist with weaknesses in cross-country comparability, one could have the reverse situation – good comparability across countries co-existing with weak comparability over time.

Thus, in the immediate and interim term, we need to accept that comparability of statistics across countries would be somewhat limited for a time and considerable effort must be dedicated to the comparability of each programme and each country's statistics over time. In other words, we must accept that global and regional aggregate statistics are somewhat crude, because the underlying national

statistics are only roughly comparable to each other. We can take advantage of the fact that programme- and country-level statistics are able to provide relatively reliable trend data. Thus, if all countries, or virtually all countries, are displaying improvements over time, we can be highly certain that the world as a whole is displaying improvement. The magnitude of global improvement could be calculated in a crude sense though not as accurately as an ideal measurement approach. However, country-level magnitudes of improvement would be reliable and certainly meaningful and useful to the citizens and governments of individual countries.

Comparability over time seems to be a latent issue not only for national initiatives but also for cross-national initiatives. It is instructive to note that even in the world's most technically-advanced cross-national assessment programmes, from time to time concerns have been raised about the comparability of national statistics over time. Challenges that deserve close attention are the strengthening of comparability over time, including a better focus on how cross-national programmes are implemented within individual countries.[6]

International education statistics would be in a healthier state if the utility of (or demand for) statistics were taken into account more effectively. This is why the UIS and its partners are not interested in collecting statistics for their own sake but in establishing a data collection system and refining those that already exist in a manner whereby a) the very process of collecting data has positive side-effects (or externalities); and b) the statistics are used constructively to bring about better education policies that advance the SDGs.

## 1.2 SDG TARGETS AND INDICATORS

The SDGs and the Education 2030 Agenda ushered in a new era of ambitions for education. Learning outcomes feature prominently in SDG 4, with five targets and six indicators calling for data on learning outcomes and skills. The reporting format of the indicators *(see **Table 1.1**)* aims to communicate two pieces of information:

a. The percentage of students/youth/adults who reach a certain level or threshold; and
b. The conditions under which the percentage can be considered comparable to the percentage reported from another country.

This requires inputs to frame the indicator:

a. What contents/skills should be measured?
b. What procedures are good enough to ensure data are comparable and of good quality?
c. A common format of reporting (scale or metrics) where all programmes could be informed with a definition of:
   - the linking methodology to the common scale in a transparent way; and
   - the definition of the threshold/minimum.

Currently, there are no common standards for a global benchmark. While data from many national learning assessments are readily available, every country sets its own objectives and standards, so the performance levels defined in these assessments may not always be consistent. This is also true with cross-national learning assessments, including international and regional learning assessments. For education systems that participated in the same cross-national learning assessments, results are comparable but not across different cross-national learning assessments and certainly not across national assessments.

The challenges of achieving consistency in global reporting go far beyond the definition of the indicators themselves. In many cases, there is no "one-stop shop" or single source of information for a specific indicator, consistent across international contexts. Even when there is agreement on the scale to be used in reporting, a harmonising process may still be necessary to ensure that programmes are comparable.

6 See Crouch and Gustafsson (2018) for a discussion of cross-sectional evidence and time-based trends.

**Table 1.1 SDG 4 targets and indicators related to learning outcomes**

| Target | Indicator | | Type | Domain | Required definitions |
|---|---|---|---|---|---|
| **4.1** By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes | 4.1.1 | Proportion of children and young people:<br><br>(a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education who achieve at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex | Global | Reading and mathematics | Minimum proficiency level<br><br>Procedural consistency |
| **4.2** By 2030, ensure that all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education | 4.2.1 | Proportion of children under 5 years of age who are developmentally on track in health, learning and psychosocial well-being, by sex | Global | Learning, socio-emotional health | Definition of "developmentally on track" |
| **4.4** By 2030, substantially increase the number of youth and adults who have relevant skills, including technical and vocational skills, for employment, decent jobs and entrepreneurship | 4.4.2 | Percentage of youth/ adults who have achieved at least a minimum level of proficiency in digital literacy skills | Thematic | Digital literacy skills | Definition of the minimum set of digital skills |
| **4.6** By 2030, ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy | 4.6.1 | Percentage of population in a given age group achieving at least a fixed level of proficiency in functional (a) literacy and (b) numeracy skills, by sex | Global | Literacy and numeracy | Definition of the fixed level of functional numeracy and literacy |
| **4.7** By 2030, ensure that all learners acquire the knowledge and skills needed to promote sustainable development, including, among others, through education for sustainable development and sustainable lifestyles, human rights, gender equality, promotion of a culture of peace and non-violence, global citizenship and appreciation of cultural diversity and of culture's contribution to sustainable development | 4.7.4 | Percentage of students by age group (or education level) showing adequate understanding of issues relating to global citizenship and sustainability | Thematic | Global citizenship and sustainability | The definition of adequate understanding and what constitutes global citizenship and sustainability |
| | 4.7.5 | Percentage of 15-year-old students showing proficiency in and knowledge of environmental science and geoscience | Thematic | Environmental science and geoscience | The definition of proficiency |

*Source: UNESCO Institute for Statistics (UIS).*

There are two extremes to consider at the time of reporting. At least in theory, greatest confidence would arise by reporting on the basis of a perfectly-equated assessment programme while, again in theory, the greatest flexibility would arise if reporting could happen with minimal alignment. Both extremes are unsatisfactory and a solution is needed on how to report with some compromise or trade-off between greatest confidence and greatest flexibility that makes use of the existing initiatives or programmes.

As a custodian agency for reporting against SDG 4, the UIS' approach is a hybrid: flexibility of reporting but with growing alignment and comparability over time, without ever necessarily reaching the extreme of a perfectly-equivalent assessment or set of assessments. This would allow any assessment programme that follows specific comparability guides, as well as quality assurance and procedural consistency, to report data in the relevant domains. This pragmatic approach implies developing tools to guide country-level work that, if complemented by capacity development activities, will ensure that the reporting of indicators drives knowledge-sharing and growth in global capacity, which will in turn use assessment programmes as levers for system improvement.

## 1.3 DATA REPORTING STRATEGY FOR SDG 4 LEARNING OUTCOME INDICATORS

Since there is no perfect solution, there is one long-term strategy for reporting with a series of short-term interim stepping stones. We can address each country's needs by adopting a portfolio approach that allows for a menu of tools for reporting and is sensitive to country specificities. The fact that the UIS is working on interim/immediate and long-term solutions also allows a high degree of practicality along the road to reaching the most "perfect" or comparable datasets.

The workflow is organized in such a way as to take two time perspectives into consideration:

### Long term

The objective is to allow the existing diversity of tools (depending on each case) to be used for reporting in the same scale based on a linking strategy that enables countries to use the same threshold as reference with a minimum set of procedures for data integrity.

### Interim/immediate

The objective is to maximise country reporting using national or cross-national initiatives that they have conducted or participated in, but that are not yet globally comparable. The UIS will footnote these criteria for short-term reporting.

### 1.3.1 Work programme

An ideal programme for reporting will have gone through three steps: conceptual framework, methodological framework and a reporting framework, as described in **Table 1.2**. Each of these contains several complex sub-steps. For various levels and types of assessment, much of this work has already been done and the focus of the work is restricted to some specific dimensions depending on the indicator.

### Conceptual framework

The design of an assessment/survey is defined by its purpose and by defining what to measure and how to measure it. The decisions made in this phase affect the possibilities of what can be done with the data collected.[7] The main questions in terms of comparing different assessment results are:

- What is the construct (for instance, reading/mathematics?) and skills/abilities measured? For example, depending on the curriculum in a country, national assessments usually have different content coverage for a given grade compared to another country.

7 Purpose, population target, test construction, domains, potential inferences, sample procedures and mode of assessing as relevant criteria for comparing the designs of assessments are key dimensions.

**Table 1.2 Key phases in an assessment programme**

| Phase | What it addresses | Main components |
|---|---|---|
| Conceptual framework | What and who to assess? | ■ Assessment/survey framework (cognitive, non cognitive and contextual)<br>■ Target population |
| Methodological framework | How to assess? | ■ Test design<br>■ Sampling frame<br>■ Operational design<br>■ Data analysis |
| Reporting framework | How to report? | ■ Defining scales<br>■ Benchmarking<br>■ Defining progress |

*Source: UNESCO Institute for Statistics (UIS).*

- What population is included? In the case of a age- or grade-based school assessment programme, it does not mean *all* children would be assessed even within the school as they might be excluded from the assessment or simply do not attend school regularly. The challenge is more serious if a large proportion of children and youth are not enrolled in school.

*Implications for global comparability*

The requirement is to define a minimum content alignment in compliance with a global content framework of reference, defining specific skills/abilities that are important for students to learn in order to function well in their communities and later in life in terms of employment.

Definitions for populations are more difficult and depend on political decisions. The sample should be at least representative of in-school children.

### Methodological framework

There are many operational issues that affect both quality and comparability. Since SDG 4 data cover many countries and include many different initiatives, it is essential to define some minimum good practices for assessment programmes to follow while respecting national authority and autonomy.

Key questions in terms of comparability are:

- Will the sample framework provide results that are valid for the population of the country? The nature of the sample is critical for the validity of the assessment programme as a measure of student learning progress at the country level, independent of any considerations of international consistency.
- Will the operational design and data generation be reliable? Robust, consistent operations and procedures are an essential part of any large-scale survey to maximise data quality and minimise the impact of procedural variation on results.

*Implications for global comparability*

There are two aspects to consider:

- Procedural alignment by complying with a minimum set of good practices on how the test was developed and how the data were collected and used in the development of the assessment.
- A variety of tools could serve to inform a given indicator. In some cases, it will be necessary to generate these tools as global public goods.

### Reporting framework

Each assessment uses different standard-setting approaches to build levels of performance so that

the scores can be classified in different categories. For education systems participating in the same cross-national learning assessments, results are comparable, but results are not comparable across different cross-national learning assessments or between national assessments.

From the point of view of reporting, there are two critical points. The first one refers to linking and the second to the definition of the minimum proficiency level.

Linking is the general term used to relate test scores on one test/form with another. Methods could be classified as equating, test calibration, projection and moderation. Others classify into equating, scale aligning and predicting. It is important to moderate differences between tests, that were designed for completely different purposes, and to express them in a way that allows some degree of comparability in the same scale. This procedure, in turn, would allow fair inferences about the subjects (countries) compared.

The second point refers to the definition of the minimum proficiency level: what is the minimum set of contents and abilities each child should know? The SDG indicators are bringing to the table a concept not yet discussed in many countries.

*Implications for global comparability*

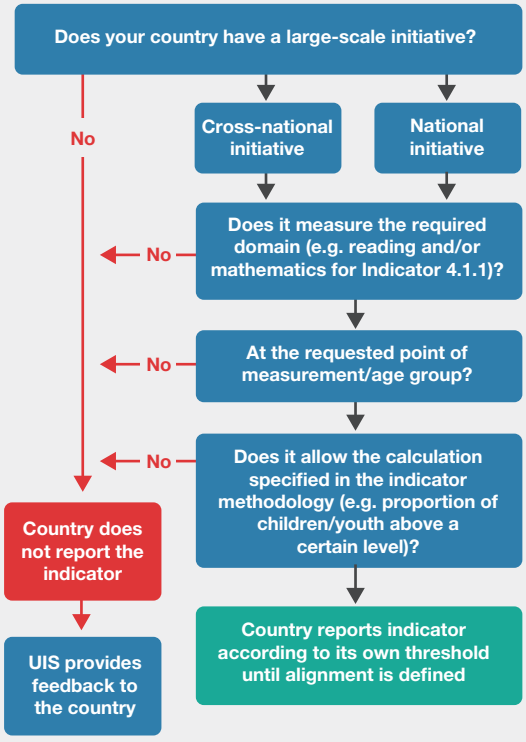
1. Alignment of results which are linked to a definition of a global point of reference as specified in each of the assessments. The solution demands flexibility and needs dialogue about critical issues – such as what each child must learn and what is the minimum.
2. Different approaches have been proposed. They all have different implications in terms of ownership, policymaking, financial costs and pedagogical implications for teachers. The way forward lies in a hybrid that embeds a portfolio approach.
3. Interim/immediate reporting starts with cross-national assessments with the comparability they permit, and all other initiatives are reported by highlighting the lack of cross-national comparability in footnotes.

## 1.4 HOW DOES A COUNTRY REPORT SDG INDICATORS TODAY?

In the first rounds of reporting, the number of caveats on comparability (limitations) is likely to outweigh the number of conditions under which cross-country comparability can be considered (possibilities). This does not detract from the value of interim reporting, recalling that the primary goal of SDG reporting is not to compare results across countries but to inform system improvement within individual countries or country groups. Over time, possibilities for international comparability may increase, but this primary purpose will remain.

Assuming that only assessment programmes with nationally-representative samples will be reported (see key considerations above), **Figure 1.1** presents the flow to be taken with footnoting beside the reported data.

**Figure 1.1 Interim reporting of SDG 4 indicators**




*Source: UNESCO Institute for Statistics (UIS).*

# 2. Reporting on Indicator 4.1.1

SDG 4 aims to promote inclusive and equitable access to quality education, as well as to the promotion of the developmental opportunities for all children and youth. This goal is operationalised as the demand to "ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes".

In particular, Indicator 4.1.1 will measure the "proportion of children and young people: (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex".

The most widely-measured areas of learning - reading and mathematics - already have a basis for global measurement, provided that national standards for primary and secondary education are used to inform local goals for the learning development of children and youth. However, this is not the case for the new global education agenda's focus on skills development in school and work to acquire the knowledge and values that promote citizenship, empathy, tolerance and sustainability.

The UIS has already published Indicator 4.1.1 proficiency statistics on its online database (for 2000 to 2017). To illustrate, 97 of 224 countries have at least one reading value for either of the two primary levels (a) and (b). These values are derived from cross-national assessment programmes and use proficiency benchmarks developed separately in each of the programmes, in other words benchmarks not intended to be comparable across the programmes.[8] This is an interim approach in the absence of a more comprehensive and country-driven system.

One of the main challenges for measurement at the global level relates to standard-setting, given the differences in context. Some of the key questions we need to answer are:

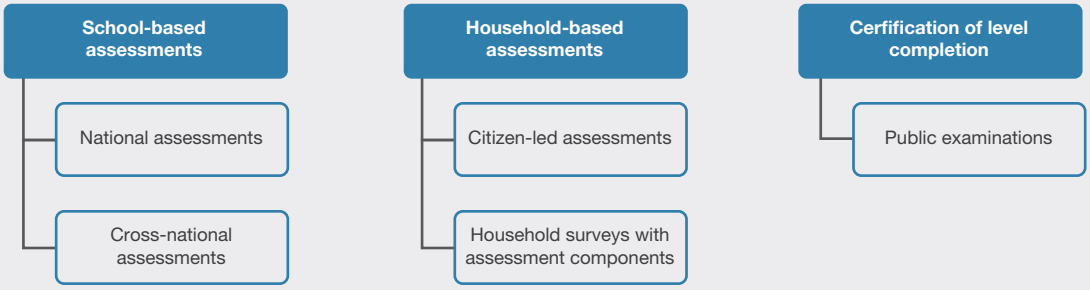
- How can the content to be evaluated be defined when it is used to align and map varied countries?
- How can contextual information be identified in the collection of background questionnaires?
- How can the minimum levels of competence and performance levels be defined?
- What kinds of guidelines are needed for data analysis and policymaking?

Alternative approaches that have been put forward differ most obviously in terms of their technical complexity, financial cost and implied comparability of national statistics. Less obvious differences relate to their sustainability over time, their impact on the politics, planning and operations of national education authorities, their ability to contribute to capacity building within countries, and their persuasive power in the media and policy debates. There are several ways in which existing proposals could be taken forward. Hybrid approaches are also possible.

This chapter aims to inform the options of a global reporting strategy for Indicator 4.1.1. Sections 2.1 and 2.2 map existing data sources, considering the coverage by point of measurement and region with special attention to school-based assessment. Section 2.3 explores the reporting options in the medium- to long-term using the framework provided in Chapter 1 and focusing on the definition of the minimum proficiency level and the linking strategy. Section 2.4 explores the progress to date, while Section 2.5 summarises the interim reporting strategy.

8 Altinok (2017) summarises these programme-specific benchmarks.

**Figure 2.1 An overview of assessment options**




*Source: UNESCO Institute for Statistics (UIS).*

## 2.1 HOW READING AND MATHEMATICS IN BASIC EDUCATION ARE MEASURED

There are numerous ways and different contexts in which reading and mathematics are measured at the national level. There is a basic distinction between assessments that are informal, formative, short or designed by teachers, inspectors and district authorities, versus formal, typically summative, longer assessments. These distinctions are important for educators because implementing short, formative assessments to monitor progress can lead to the development of more complete summative assessments.

Large-scale assessments can be divided into two categories: school-based and household surveys *(see **Figure 2.1**)*.

School-based assessments include two types:

- National assessments (or, in principle, sub-national assessments as may occur in decentralised or federal countries) designed to measure specific learning outcomes at a particular age or grade that are considered relevant for national policymakers; and
- Cross-national initiatives (either regional or international) administered in a number of countries, based on a commonly agreed upon framework, following similar procedures yielding comparable data on learning outcomes.

Household-based learning assessments can be used to target populations that may or may not be enrolled in or attend school. They include any household surveys that include an assessment component in their data collection.

A particular case within this last category are citizen-led assessments originating in non-governmental organizations or think tanks and are meant to exert accountability pressure on governments and to engage citizens. There are various reasons why these assessments are household-based. A primary reason is that such assessments can "capture" the skills of children regardless of whether they are enrolled in school or not *(see PAL Network)*.

Both household-based surveys and school-based assessments collect background information that add context to data on learning outcomes. By including children and young people in and out of school, household-based surveys provide information on families and enabling environments. School-based assessments provide system-level information on classroom and school environments and sometimes gather information about the home environment either via a parent or via child recall. Together, school-based assessments and household-based learning assessments help to provide a snapshot of how

## Box 2.1 Where and how to find SDG 4 data

- The *Quick Guide to Education Indicators for SDG 4* describes the process of developing and producing the global monitoring indicators while explaining how they can be interpreted and used. This is a hands-on, step-by-step guide for anyone who is working on gathering or analysing education data.
- The *SDG 4 Data Book: Global Education Indicators 2018* ensures that readers have the latest available data for the global monitoring indicators at their fingertips, and will be regularly updated.
- The SDG 4 Data Explorer, displays data by country, region or year; by data source; and by sex, location and wealth. It allows users to explore the measures of equality that are crucial for the achievement of SDG 4.
- UIS.Stat is the world's most comprehensive database on education. It enables users to search and extract data from across UIS's many databases.
- The SDG 4 database contains data on key indicators needed for global monitoring, including data on **learning outcomes**. It presents the assessment undertaken by each country as well as the share of children who reached minimum proficiency levels in reading and mathematics.

*Source: UNESCO Institute for Statistics (UIS).*

children and youth around the world are learning. However, the results from these different types of assessments cannot, for now, be legitimately compared internationally or internally within a country.

Public examinations have high-stakes that apply to all individuals at certain points in the grade structure of an education system. They serve to select students for continuing education programmes or to certify attainment of a certain qualification.

## 2.2 REPORTING ON INDICATOR 4.1.1

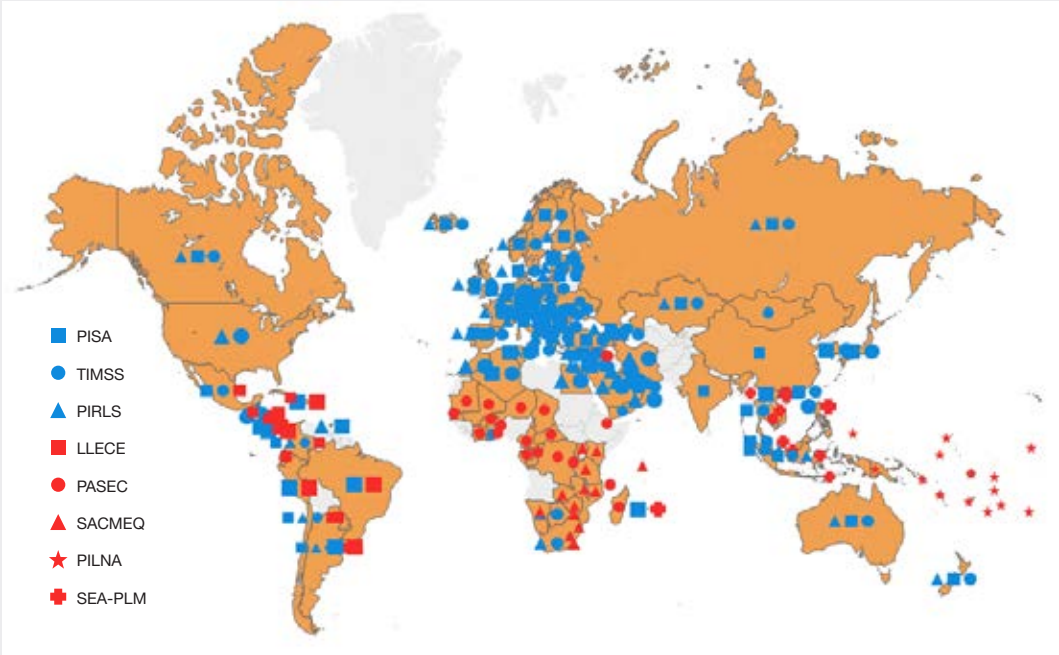
All cross-national assessments – both global and regional – and national assessments could be used to inform Indicator 4.1.1. Naturally, the cross-national assessments (aiming at both global and regional coverage) have been first in line to report for SDG 4.1.1 as they are designed for cross-national comparisons, measure common subject areas or assessment domains (minimum core denominator) and are expressed on a common scale. Unfortunately many regions do not have a regional assessment, nor have the countries joined any cross-national initiative. This represents a challenge if the options are to be restricted to cross-national assessments.

### 2.2.1 How global and regional assessments are distributed globally

**Figures 2.2** and **2.3** map the current distribution of assessments by category. In terms of subjects assessed, reading and mathematics are the most common areas of study. As previously explained, all cross-national and national assessments could be used to inform Indicator 4.1.1. In addition, household-survey based assessments, in general those measuring foundational skills such as Multiple Indicator Cluster Surveys (MICS), Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) can be used as well.[9] According to UIS estimates, 80% of countries have conducted a national learning assessment or participated in a cross-national initiative in the last five years (UIS, 2016). This represents a significant increase in the number of student assessments undertaken globally over the past decade. This increase is largely due to the growing number of countries interested in monitoring their progress in a regional context, leading to a rapid growth of regional assessments during this period. However, due to differences

9 See Treviño and Ordenes, 2017.

**Figure 2.2 School-based assessment**




**Note:** Areas shaded in orange correspond to the existence of national assessments.
*Source: UNESCO Institute for Statistics (UIS).*

**Figure 2.3 Foundational skills assessments – countries implementing household-based assessments in basic education**


*Source: UNESCO Institute for Statistics (UIS).*

**Table 2.1 Options for reporting on Indicator 4.1.1**

| | School-based | | Population-based |
|---|---|---|---|
| | Cross-national | National | |
| Grade 2 or 3 | LLECE<br>PASEC<br>TIMSS<br>PIRLS | Yes | MICS6<br>EGRA/EGMA<br>PAL Network |
| End of primary education | LLECE<br>PASEC<br>SACMEQ<br>PILNA<br>SEA-PLM<br>TIMSS<br>PIRLS | Yes | PAL Network |
| End of lower secondary education | TIMSS<br>PISA<br>PISA-D | Yes | Young Lives |

**Notes:**
EGMA: Early Grade Mathematics Assessment
EGRA: Early Grade Reading Assessment
LLECE: Latin American Laboratory for Assessment of the Quality of Education
MICS6: Multiple Indicator Cluster Surveys, Round 6
PAL: People's Action for Learning
PASEC: *Programme d'analyse des systems éducatifs de la CONFEMEN* (Programme of Analysis of Education Systems of CONFEMEN)
PILNA: Pacific Islands Literacy and Numeracy Assessment
PIRLS: Progress in International Reading Literacy Study
PISA: Programme for International Student Assessment
PISA-D: Programme for International Student Assessment for Development
SACMEQ: Southern and Eastern Africa Consortium for Monitoring Educational Quality
SEA-PLM: Southeast Asia Primary Learning Metrics
TIMSS: Trends in International Mathematics and Science Study
*Source: UNESCO Institute for Statistics (UIS).*

in the measurement constructs and frameworks, these assessments are not always comparable across countries and many technical challenges remain. Thus, it is difficult at this stage to compare countries in terms of learning achievement across regions due to the lack of comparability using the same scale.

Knowing what currently exists in countries in terms of assessment systems is important for charting a way forward for Indicator 4.1.1 and finding a feasible, cost-effective way of reporting. Indicator 4.1.1.a – that is, for early grades – is classified in Tier III, while Indicators 4.1.1.b and c – that is, primary and lower secondary levels – are classified in Tier II. Expanding the linking and reporting options could be the way forward to upgrading the sections of the indicator in Tiers II and III. [10]

## 2.2.2 Understanding the current configuration of school-based assessments

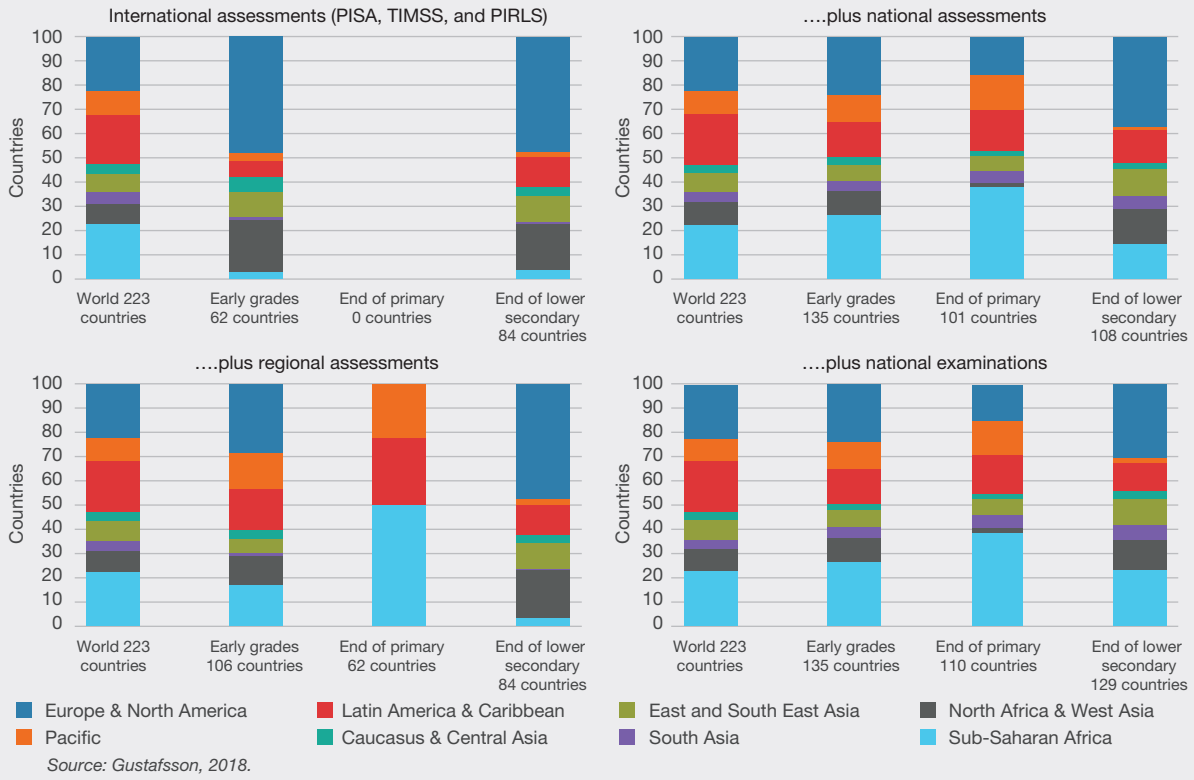
For the purpose of analysis, and given their relevance for each of the three educational levels of Indicator 4.1.1, we will limit our discussion to school-based assessments. It is useful to consider four types of school-based assessments, each offering specific opportunities and challenges: i) the three large international programmes (PISA, TIMSS and PIRLS); ii) the five regional cross-national assessments (LLECE, PASEC, PILNA, SACMEQ and SEA-PLM); iii) national assessments for monitoring purposes (either sample- or census-based); and iv) national examinations for certification or selection purposes.

**Figures 2.4** and **2.5** focus on differentiating coverage in terms of the three educational levels and the four types of assessments with two criteria of coverage (number of countries and population). Figure 2.4 focuses on the number of countries by region and level regardless of the size of the countries. On the other hand, Figure 2.5 refers to population coverage. As the SDG indicators follow a tier classification - that means all regions need a reasonable coverage - the analysis is presented in terms of regions and not as a particular assessment.

Figure 2.4 indicates that the three global assessments provide the best coverage at the lower secondary level if only international assessments are considered. Yet even here, fewer than one-half of the world's countries are covered, though participating countries represent 76% of the world's population (as shown in Figure 2.5). Adding the five regional programmes expands coverage for the end of primary education.

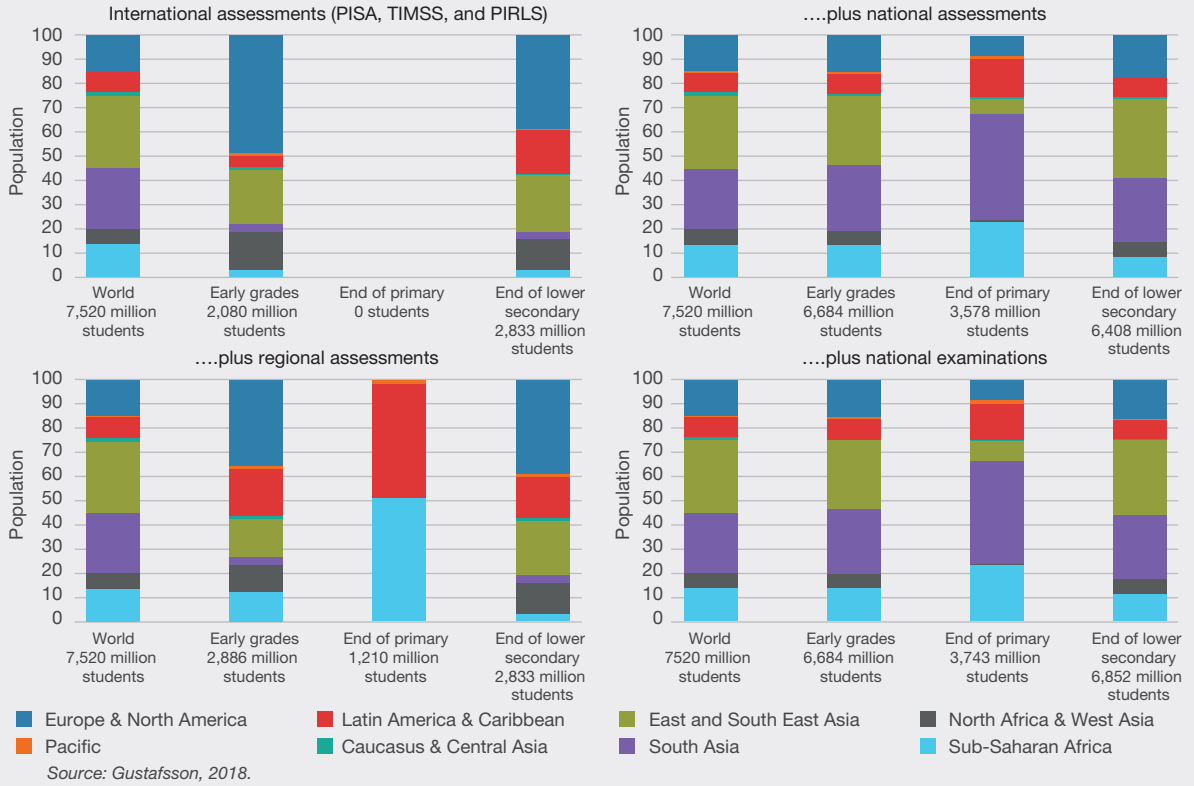
10 Tier 2: Indicator is conceptually clear, internationally-established methodology and standards are available, but data are not regularly produced by countries.
Tier 3: No internationally-established methodology or standards are yet available for the indicator, but methodology/standards are being (or will be) developed or tested.

## Figure 2.4 Country coverage by type of assessment




*Source: Gustafsson, 2018.*

## Figure 2.5 Population coverage by type of assessment and region




*Source: Gustafsson, 2018.*

Very large gains are visible following the addition of national assessments, though it is not clear if all meet the minimum procedural quality or have enough alignment in terms of content coverage to make them comparable. This would be a general problem unless some minimum content coverage is developed. However, they could still report Indicator 4.1.1 with footnotes referring to the shortcomings that could serve as a warning in terms of comparability.

Catering for the examinations of countries adds relatively little coverage, though at the lower secondary level the 7 percentage point gain (86% to 93%) is substantial.

### 2.2.3 Why is it relevant to define the minimum level?

The Education 2030 Framework for Action commits all countries to establish benchmarks for measuring progress towards SDG 4 targets. In response, the UIS has proposed the development of scales that describe the progression of learning skills and thereby help countries to identify and agree on the benchmarks needed to define minimum proficiency levels for reporting purposes *(see Section 2.3.2)*.

It is important to recognise that several years will be required to resolve all of the methodological and political issues needed to report on SDG indicators on the same scale. The challenges are primarily due to the fact that learning assessment initiatives use different definitions of performance levels, while discussions continue on an interim reporting strategy. The UIS has published a database that links different assessments to the same scale and to report on Indicator 4.1.1 using two alternative benchmarks.

The two benchmarks belong to two different assessments that reflect the contexts of countries with different income levels. For example, SACMEQ is a regional survey used to assess students at the end of primary school. The decision was therefore made to use the SACMEQ benchmark (referred to as the basic proficiency level) for reading and mathematics at the primary level for all countries in the database *(see **Box 2.2**)*.

In addition, the database includes results using the minimum proficiency level defined by the International Association for Evaluation of Educational Achievement (IEA) for the Progress in International Reading Literacy Study (PIRLS) and Trends in International Mathematics and

#### Box 2.2 What are children expected to know at the primary level?

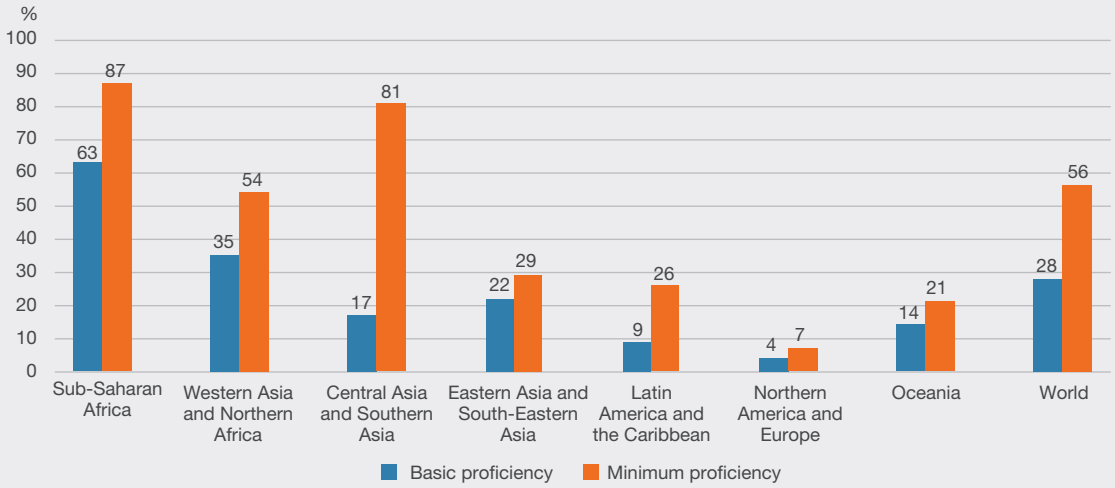
According to the SACMEQ benchmarks, children in Grade 6 who have achieved the minimum proficiency level in reading can "interpret meaning (by matching words and phrases completing a sentence, matching adjacent words) in a short and simple text by reading forwards or backwards" (SACMEQ III).

In mathematics, students can "translate verbal information (presented in a sentence, simple graph or table using one arithmetic operation) in several repeated steps". Moreover, he/she "translates graphical information into fractions, interprets place value of whole numbers up to thousands and interprets simple common everyday units of measurement" (Hungi et al., 2010).

The IEA benchmarks used in PIRLS and TIMSS are more demanding. For example, "when reading Informational Texts, students can locate and reproduce explicitly stated information that is at the beginning of the text" (Mullis et al., 2012). For mathematics, "students can add and subtract whole numbers. They have some recognition of parallel and perpendicular lines, familiar geometric shapes and coordinate maps. They can read and complete simple bar graphs and tables" (Mullis et al., 2016b).

*Source: UIS, 2017g.*

**Figure 2.6 Proportion of students not reaching the basic and minimum proficiency levels in reading by SDG region**




**Note:** The minimum level is higher than the basic level of proficiency. Hence, more children do not achieve the minimum proficiency level than do achieve the basic level.
*Source: UNESCO Institute for Statistics (UIS).*

Science Study (TIMSS). Both of these international assessments have global coverage primarily involving middle- and high-income countries.

**Figure 2.6** shows the percentage of primary and lower secondary students not achieving the basic proficiency level and the minimum proficiency level. The minimum proficiency level is more difficult and requires a higher level of skills and concepts, which explains why fewer students are achieving it. Hence, less children achieve the minimum proficiency level than the basic level.

It is also important to note the variation in rates between regions. The change in percentage of students below the basic and the minimum proficiency level is not linear. Linearity could occur if there were a similar distribution of pupils for all possible scores between countries. A high proportion of students concentrated around the basic proficiency level implies that a minor change in the levels of the threshold to the minimum proficiency level will produce a dramatic reduction in the proportion of children who reach minimum proficiency levels. There are regions with a high proportion of children with very basic sets of skills for whom the minimum proficiency level is too high a bar. This explains why such a high proportion is not reaching the benchmark.

The differences in the results highlight the need to accelerate discussions on benchmarks. Is it possible to define appropriate benchmarks for all? There is a clear need to define concepts as well as to examine the feasibility and utility of setting benchmarks at different levels of monitoring. Both the technical and political aspects of the process must be taken into account in these discussions.

## 2.3 A FRAMEWORK FOR REPORTING INDICATOR 4.1.1

The reporting format aims to communicate two pieces of information:

1. The percentage of students meeting minimum proficiency standards for the relevant domain (mathematics and reading) and measurement point (early grades, end of primary education and end of lower secondary education); and
2. The conditions under which the percentage can be considered comparable to the percentage reported by another country.

This requires decisions on the following:

1. What content should be measured and what is the percentage of coverage to be covered by a given assessment to make it comparable with others?
2. What procedures are sufficient to assure the quality of the data collected?
3. A proficiency scale to ensure comparability.
4. The definition of the set of skills/contents defined as "minimum".
5. A method of linking assessment programmes to the scale.

### 2.3.1 Challenges

The challenges of achieving consistency in global reporting go far beyond the definition of the indicators themselves. In many cases, there is no "one-stop shop" or single source of information for a specific indicator that is consistent across international contexts. Even when there is agreement on the metric to be used in reporting, a harmonising process may still be necessary to ensure that coverage of the data is consistent.

This entails creating common methodologies to ensure comparability of the data that currently exist, as well as promoting the development of new assessments to collect any data that are not yet available. In political terms, the challenge will come when the leaders need to agree on a "minimum proficiency level" that is conceptually adequate and relevant for all countries.

A study conducted by Treviño and Ordenes (2017) sets the stage by exploring the commonalities and differences between regional and international assessments, with the objective of understanding the challenges and options in terms of reporting on Indicator 4.1.1.

The analysis suggests that:

- All the different approaches to measuring Indicator 4.1.1 have advantages and shortcomings in relation to technical issues and feasibility.
- It is necessary to create political agreement and advance the technical sphere to define the minimum level of competency in reading and mathematics.
- It is also necessary to approach procedural consistency so that a minimum level of data quality is established, given the heterogeneity among assessment programmes.
- Four strategies for reporting Indicator 4.1.1 are possible, including a new, unique SDG 4 test.
- The alternative of developing a specific instrument with a clear definition of the minimal level of competency may ensure high levels of comparability of results and avoid technical critique, but loses flexibility and is politically difficult to sell.

### 2.3.2 Reporting consistency: The UIS work flow

Since 2016, the UIS has been working with partners and discussing options through GAML *(see **Box 2.3**)*.

**Table 2.2** contextualises all the work underway to report on SDG 4. Column 3 highlights UIS work to help fill gaps.

The objective is to define the criteria and generate the tools that could serve as:

- **Reference points:**

The content, procedural and reporting alignment provide a common language and approach to the development of assessment contents (for mathematics and reading), minimum procedural practices and reporting that will ensure comparable monitoring of progress towards Indicator 4.1.1.

- **Transparency tools:**

The adoption of common minimum coverage practices and a reporting framework could make comparisons more transparent across countries and regions.

## Box 2.3 Global Alliance to Monitor Learning

The Global Alliance to Monitor Learning (GAML) is a multi-stakeholder initiative aimed at addressing measurement challenges based on consensus and collective action in the learning assessment arena, while improving coordination among actors.

GAML brings together UN Member States, international technical expertise and a full range of implementation partners - donors, civil society, UN agencies and the private sector - to improve learning assessment globally. Through participation in GAML, all interested stakeholders are invited to help influence the monitoring of learning outcomes for SDG 4 and the Education 2030 goals.

GAML operates through task forces which have been established to address technical issues and provide practical guidance for countries on how to monitor progress towards SDG 4. The task forces make recommendations to GAML on the framework for all global and thematic indicators related to learning and skills acquisition, tools to align national and cross-national assessments into a universal reporting scale for comparability, as well as mechanisms to validate assessment data to ensure quality and comparability.

*Source: UNESCO Institute for Statistics (UIS).*

### Table 2.2 Summary of processes and the focus of GAML

| Phase/tool | What it addresses | Focus of UIS work | Products generated/ tools for countries | Status |
|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) |
| Conceptual framework | What to assess? - Concept<br><br>Who to assess? – Population: in and out of school?<br><br>What contextual information to collect? | Global Content Framework | Global Content Framework (GCF) to serve as reference | Finalised |
| | | | Content Alignment Tool (CAT) | Draft for approval |
| | | | Online platform for CAT | Draft for approval |
| Methodological framework | What are the procedures for data integrity? | Procedural alignment | Manual of good practice | Finalised |
| | | | Quick guides to support implementation in countries (3) | Finalised |
| | | | Procedural Alignment Tool<br><br>Online platform | Finalised |
| Reporting framework | What format to report?<br><br>What is the minimum level?<br><br>How to link or “harmonise”? | Proficiency framework and minimum level<br><br>Linking strategies<br><br>Interim reporting | Scale and definition of minimum proficiency level | Draft for adoption |
| | | | A linking strategy portfolio | Draft for adoption |
| | | | An interim reporting strategy | Finalised |

*Source: UNESCO Institute for Statistics (UIS).*

### Global Content Framework

This section describes in more detail the work that needs to be done or is underway for Row 1, Column 4 in Table 2.2.

*Why?*
Assessment programmes differ in the conceptual frameworks that are used to develop their overall assessment framework. For example, depending on the curriculum in a country, national assessments usually have different content coverage for a given grade. Furthermore, even domains can be defined differently. In some cases, programmes assess different skills, sometimes they use different content to assess the same domain, and sometimes they do both differently, even for the same grade.

To assess the degree of alignment among various assessments and to begin to lay out the basis for a global comparison, the UIS and the International Bureau of Education (IBE-UNESCO) have jointly developed a Global Content Framework (GCF) for each of the domains of mathematics and reading (upper right-hand cell in Table 2.2).

*Scope of UIS work*
a. To define the minimum common set of contents and skills that should be taught and assessed in each of the points (Grade 2 or 3, at the end of primary education and at the end of lower secondary education) of measurement that the indicator requires.
b. To facilitate the tools for countries to assess the alignment of content.

### Procedural alignment

This section describes in more detail the work that needs to be done, or is being done, for Row 2, Column 4, in Table 2.2.

*Why?*
Assessment implementation faces many methodological decisions that are not identical, tests can be built in different formats, the sampling decisions are not identical, etc. There is no need for identical procedures and format, but there is a need for a minimum set of procedures so that data integrity is protected and results are robust as well as reasonably comparable for any given country over time (most important) and across countries at any given point in time (less important but still relevant).

Robust, consistent operations and procedures are an essential part of any large-scale assessment to maximise data quality and minimise the impact of procedural variation of results. Examples of procedural standards may be found in all large-scale international assessments, and for many large-scale assessments at the regional level, where the goal is to establish procedural consistency across international contexts. Many national assessments also set out clear procedural guidelines to support consistency in their operationalisation.

*Scope of UIS work*
a. To define minimum procedural practices that ensure integrity in the data-generating process through guidance on good practices.
b. To generate a tool for countries to assess their alignment (Table 2.2, Column 4).

### Reporting

This section describes in more detail the work that needs to be done or is underway for Row 3, Column 4, in Table 2.2.

*Why?*
Assessment programmes typically report using different scales. Analysis of results therefore remains limited to a particular test, linked to one methodology and one scale. Although some convergence takes place between international and regional assessments, it is still difficult to situate assessments in a common reference continuum of learning outcomes for each level and domain.

The most important issue in the definition of the scales is the proficiency benchmarks or levels embedded within the numerical scale and their cut-points on that numerical scale. These benchmarks are typically associated with proficiency level descriptors (PLDs), which describe in some detail the skills that are typical of students at any given cut-point in the scale. Typically, an overarching policy statement or policy definition gives meaning to the succession of cut-scores and the proficiency levels but most importantly for defining what constitutes a *minimum* (which is what Indicator 4.1.1. calls for) proficiency level that has reference to the content.[11]

Currently, there is no common standard as a global reference. While data from many national learning assessments are available now, every country sets its own standards so that the performance levels defined in these assessments may not always be consistent. This is also true of cross-national learning assessments, including international and regional learning assessments. For education systems that have participated in the same cross-national learning assessments, results are comparable but not across different cross-national learning assessments and certainly not across national assessments.

*Scope of UIS work*

a. To define a scale to locate all the learning assessment programmes.
b. To establish the linking strategy to that scale.

### How to define the minimum proficiency level?

The definition of a minimum proficiency level has both political and technical implications. This is a critical definition when applicable to different contexts and situations. A simple example helps to illustrate this discussion: according to the SACMEQ benchmarks, children in Grade 6 who have achieved the minimum proficiency level in reading can "interpret meaning (by matching words and phrases completing a sentence, matching adjacent words) in a short and simple text by reading forwards or backwards" (SACMEQ III), while for IEA's PIRLS, the minimum level is defined as "when reading Informational Texts, students can locate and reproduce explicitly stated information that is at the beginning of the text" (Mullis et al., 2012).

The UIS has taken a pragmatic approach that consists of using the existing set of proficiency levels that are widely used (and validated) by countries participating in a global or international assessment as part of the process of reporting. The UIS is basically:

a. Mapping all proficiency levels with their descriptors;
b. Aligning in a continuum from lower to higher level;
c. Mapping the points in each assessment that define the minimum proficiency level and its policy descriptors;
d. Based on previous mapping, defining a minimum level and building consensus; and
e. Once steps 1 to 4 are finished, defining "preliminary" PLDs.

**Figure 2.7** provides an example in mathematics testing.

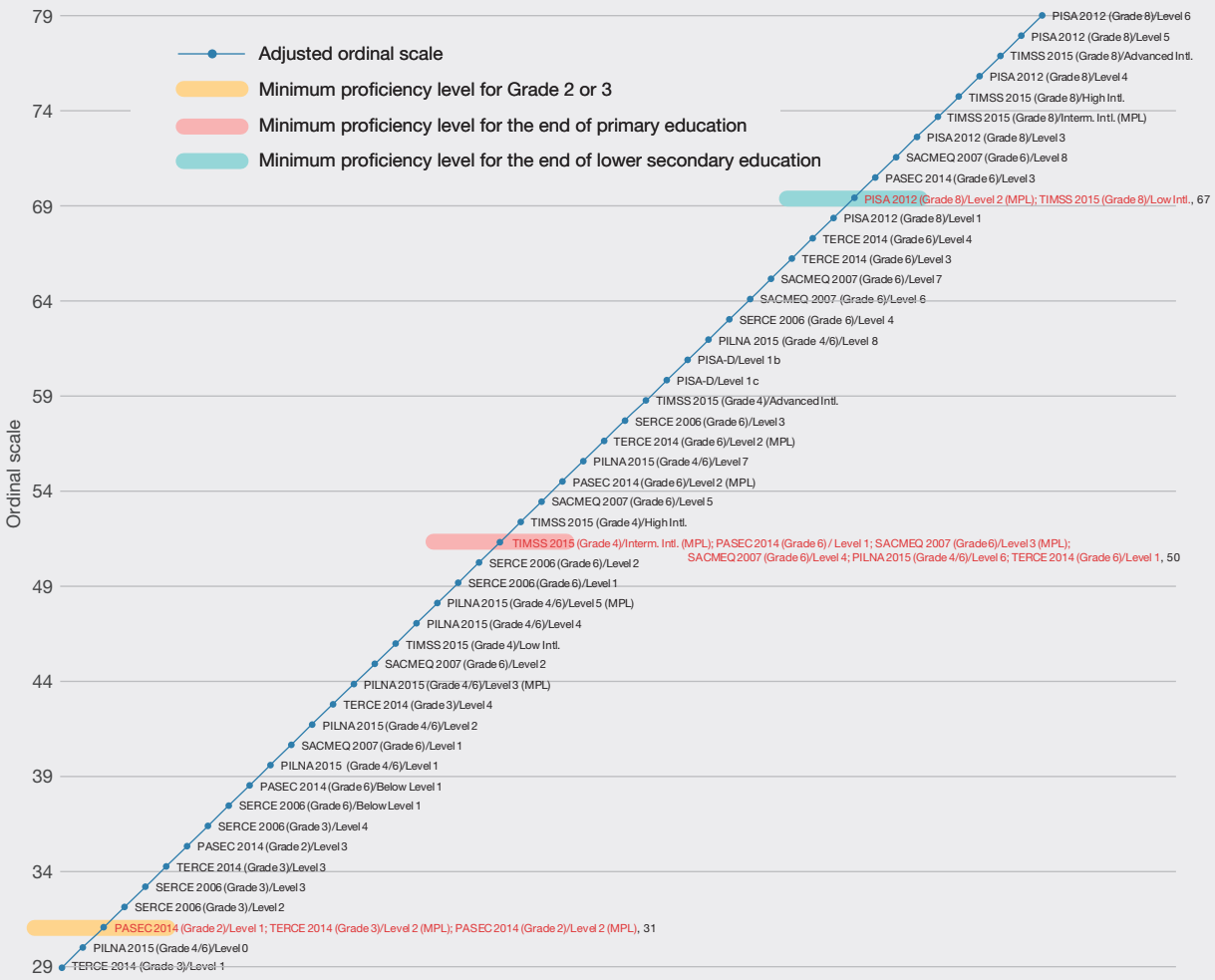
A technical meeting with partners in September 2018 found consensus about the minimum proficiency level definition as reflected in **Table 2.3**. The next step will be to provide a general description and details of tasks and examples of items from different tests. For example, the descriptor for the global minimum proficiency level for Grade 3 corresponds to Level 4 of PASEC and Level 1 of TERCE. So despite their apparent differences, the global minimum proficiency level shows the correspondence between these two regional tests. This approach applies to the other assessments listed below.

Thus for a given country, various minimum proficiency levels could co-exist. The first is the national level with its objective based on national policies and national curriculum. The second is a regional reference and regional assessments if they exist and a global reference as agreed upon.

11 Taking from the NAEP on policy statement: "Policy definitions are general statements to give meaning to the levels".

## Figure 2.7 Proficiency scales in mathematics according to current PLDs

Mathematics proficiency scale and minimum proficiency levels




**Notes:** Proficiency levels below the scale: All proficiency levels from ASER 2017, EGMA, Uwezo and UNICEF MICS6; PASEC 2014 Grade 2 (below level 1); SERCE 2006 Grade 3 (Level 1 and below Level 1). MPL: Minimum proficiency level as defined by each assessment.
*Source: UNESCO Institute for Statistics (UIS).*

### Linking to the proficiency scale

The linking of a national and/or regional assessment to the global definitions will require in-depth enquiry into the assessment items. Linking is the general term used to relate test scores from one test/form to another test/form. Different researchers have proposed different approaches. But overall, linking is about moderating differences between tests that were designed for completely different purposes to express them in the same scale in a way that allows some degree of comparability that, in turn, allows fair inferences about the subjects (countries) compared. The process of making different tests comparable is generally referred to as “moderation”.

Statistical moderation utilises the score distribution of two assessments to construct concordance tables mapping the scores on two tests that do not measure the same constructs. Methods such as calibration

**Table 2.3 Minimum proficiency level alignment for mathematics**

| Educational level | Descriptor | Assessment PLDs that align with the descriptor | Minimum proficiency level in the assessment |
|---|---|---|---|
| Grades 8 and 9 | Students demonstrate skills in computation, application problems, matching tables and graphs, and making use of algebraic representations. | PISA 2015, Level 2 | Level 2 |
| | | TIMSS 2015, Low International | Intermediate international |
| Grades 4 and 6 | Students demonstrate skills in number sense and computation, basic measurement, reading, interpreting, and constructing graphs, spatial orientation, and number patterns. | SACMEQ 2007, Level 3 | Level 3 |
| | | SACMEQ 2007, Level 4 | |
| | | PASEC 2014, Level 1 | Level 2 |
| | | PILNA 2015, Level 6 | Level 5 |
| | | TERCE 2014, Level 1 | Level 2 |
| | | TIMSS 2015 Intermediate international benchmark | Intermediate international |
| Grade 2 or 3 | Students demonstrate skills in number sense and computation, shape recognition and spatial orientation. | TERCE 2014, Level 2 | Level 2 |
| | | PASEC 2014, Level 1 | Level 2 |
| | | PASEC 2014, Level 2 | |

*Source: UNESCO Institute for Statistics (UIS).*

(putting items and persons on one test form onto the same scale and setting a reference point) and equating (setting up a common scale for different tests, removing unintended differences in test form difficulties and setting up a common scale) refer to alternative ways of linking. It is important to keep in mind that the strength of the linking depends on the degree of similarity between inferences, constructs, populations and measurement conditions.

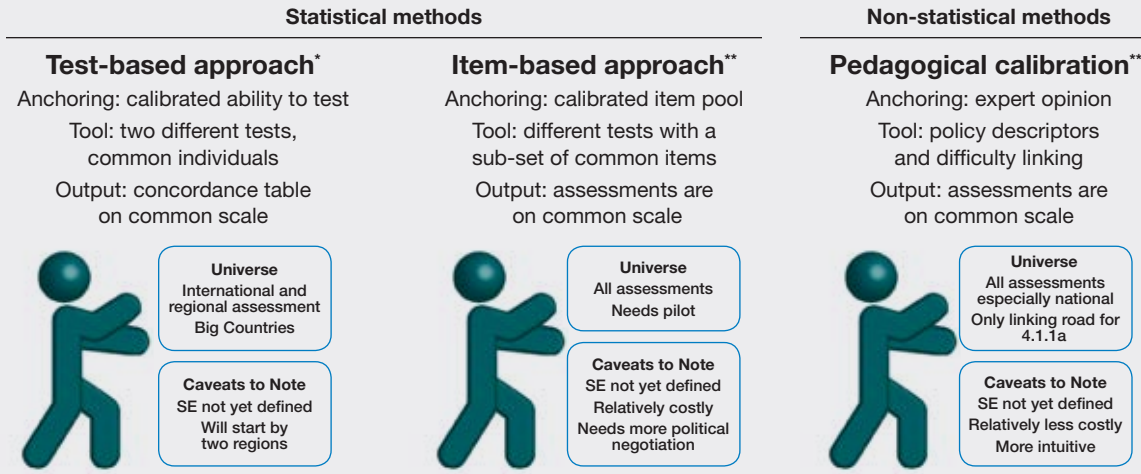
Non-statistical moderation has the same objective as statistical moderation, but the concordance table of comparable scores are obtained by matching tests scores by subjective judgement of experts. In general, this is described as "social moderation" or pedagogical recalibration because it uses judgement to match levels of performance with different assessments directly. Thus, social moderation calls for direct judgement about the comparability of performance levels between different assessments.

As statistical moderation is based on comparability at a certain point in time of certain items or individuals, social moderation comparability comes from the opinion of a group of people as the social moderators, rather than a set of students or items at a certain point in time. Nobody can solve all of the uncertainty involved in these choices (items, students, moderators) and there is always some subjectivity.

However, social moderation could serve to define (and establish) broad standards for the knowledge and skills that students have to achieve. It can also be used to monitor performance and understand the meaning of a minimum level that students are expected to know and be able to do in relation to grade-appropriate content. This lies at the heart of the curricular definitions in any country.

Moderation or linking is not an application of the principles of statistical inference but a way to specify the rules of the game. Establishing the rules of the

**Figure 2.8 Innovative solutions to generate comparable data for Indicator 4.1.1**




**Notes:** The UIS Proficiency Scale is the reference scale for reporting on Indicator 4.1.1, after all assessments are put on a common scale.
* Test-based approach: Common individuals meaning representative individuals of similar characteristics are presented with two different tests.
** Item-based approach: Common items different tests taken by different individuals. Tests will be put on common scale once embed the calibrated items from the item pool.
*** Pedagogical calibration approach: Use content/context experts with relevant experience in country to generate consensus on the alignment of national assessment to a Proficient Scale taking into account constructs and difficulties of the items. No extra field work required.

*Source: UNESCO Institute for Statistics (UIS).*

game would help to build agreement on a way of comparing students who differ quantitatively but does not provide information about tests that are not built to measure the same construct. Consensual processes and experts are the way forward.

The proposals for linking to a common scale are clearly not mutually exclusive. The proposals described here are in fact a combination of approaches that aim to establish some rules for comparing students, youth and adults. Alternative strategies to achieve comparability and assessing their effectiveness and efficiency are a matter of proof.

*Scope of UIS work*

a. To define a set of cost-efficient linking strategies to maximise reporting.
b. To define an immediate/interim solution to reporting.

The UIS has taken a portfolio approach that includes two broad sets of possibilities: the non-statistical approach and the statistical approach that differ in the degree that they rely on "hard" psychometric evidence to define comparability. **Figure 2.8** summarises the options below.

### Strategy 1. The non-statistical approach: Pedagogically-informed recalibration of existing data

The approach involves using the proposed proficiency framework that describes the range of competencies that children and youth have at each level to locate proficiency levels from alternative assessment programmes based on the performance level descriptors. The approach is referred to as social moderation because linking is guided by expert judgement. This proposal would allow the expansion of coverage in terms of educational systems reporting for SDG 4. For instance, coverage at the primary level would double, in terms of the population-weighted world, if national assessments were included.

**Table 2.4 Relationship between linking strategies and coverage of assessment type**

| | Statistical linking | | | Judgmental linking |
|---|---|---|---|---|
| | Recalibration through parallel tests | Psychometrically-informed recalibration | Statistical recalibration of existing data | Pedagogically-informed recalibration |
| PISA, TIMSS and PIRLS | Will be used | Could be used | Yes | Yes |
| Regional cross-national assessments | Will be used | Could be used | Yes | Yes |
| National assessments | Could be used | Could be used | Not clear how | Yes |
| National examinations | – | – | Not clear how | To be used |

*Source: Gustaffson, 2018.*

### Strategy 2. The statistical approach

*2.a. Psychometrically-informed recalibration based on common items*

- Implies the use of common items in different assessment programmes.
- One version has been proposed by the Australian Council for Educational Research (ACER) as part of an overall proposal of progression in learning but options are not exhausted.[12]
- Has proven to face many difficulties in implementation from technical and political perspectives.

*2.b. Recalibration by running a parallel test on a representative sample of students*

- The IEA outlines the "Rosetta Stone" solution (*see Annex 2*) that deals only with the primary level and allows two assessments (one international and the other regional) to be expressed on the same scale. Concretely, the proposal states that sub-samples of students in three to five countries per programme would write not just the regional tests but also IEA's test.
- This would produce a "concordance table" with all countries participating and not participating in the same scale based on psychometric modelling.[13] The table is not the reporting scale itself but facilitates it by expressing a larger number of countries in the same scale.

*2.c. Recalibration of existing data*

- This approach relies largely on statistical adjustments[14] taking advantage of the fact that some countries, referred to as "doubloon countries", participate in more than one cross-national programme. Using several such overlaps has allowed for the identification of roughly-comparable proficiency thresholds. It could serve as a double-check, but its political buy-in is unlikely.

*Weighing options*

The efforts described in **Table 2.4** should be taken more as complementary routes than as alternative options in order to minimise risk if some of the approaches prove to be too costly, the margin of error is too high, politically-unfeasible or a combination of these issues. The approaches help to build a sustainable reporting strategy where it is easier to see stepping stones between **Strategy 1** and **Strategy 2a** and complementarity between

12 Note that the reference scale is built from items from various assessments.

13 For countries the option is to either participate in a regional or global programme (something that might be difficult or not possible if the region does not have a regional initiative).

14 See Altinok, 2017.

**Table 2.5 How interim reporting is structured**

| | School-based | | Population-based | Grade to be assessed |
|---|---|---|---|---|
| | Cross-national | National | | |
| Grade 2 or 3 | LLECE<br>PASEC<br>TIMSS<br>PIRLS | Yes | MICS6<br>EGRA/EGMA<br>PAL Network | 2 or 3 |
| End of primary education | LLECE<br>PASEC<br>SACMEQ<br>PILNA<br>SEA-PLM<br>TIMSS<br>PIRLS | Yes | PAL Network | Plus or minus one year of last year of primary education according to ISCED level in a country |
| End of lower secondary education | TIMSS<br>PISA<br>PISA-D | Yes | Young Lives | Plus **two** minus one grade of last year of lower secondary education according to ISCED level in a country |
| Definition of minimum level until 2018 release | The ones defined by each assessment by point of measurement and domain | | | |
| Definition of minimum level from 2019 | According to alignment as adopted by GAML and TCG | | | |
| Grade for end of primary and end of lower secondary education | As defined by the ISCED level of each country | | | |
| Validation | Send from the UIS for country approval | | | |

**Notes:** * TIMSS/PIRLS Grade 4: These results are allocated to the end of primary education when, according to the ISCED levels in a given country, there are four grades in primary education. When primary education has more than four grades, they are allocated to Grade 2 or 3.
** The UIS advises to complement this indicator with the indicator on out of school children.
*Source: UNESCO Institute for Statistics (UIS)*

**Strategy 2b** and **Strategy 1**, such as the Rosetta Stone which needs to be expressed in a proficiency framework. Strategy 2c could be potentially used as a check to compare statistics based on national assessments (Treviño and Ordenes, 2017).[15]

## 2.4 INTERIM REPORTING STRATEGY FOR INDICATOR 4.1.1

The UIS has defined an interim reporting period with a strategy that encourages maximum reporting, makes full use of available information and acknowledges the shortcomings of data by footnoting, while releasing and providing the standards to improve quality and reporting in the same scale.[16]

This means that the minimum proficiency level will be reported according to what is informed by each assessment, without having been expressed in the same scale, as summarised in **Table 2.5**, and would follow the flow as described in **Figure 2.9**. Over time, there would be possibilities for international comparability and better quality data.

15 A third strategy could be a new test that everybody takes for reporting using a common comparable tool but this is neither politically-feasible nor cost-efficient so it has not been pursued.

16 This does not detract from the value of interim reporting, recalling that the primary goal of Indicator 4.1.1 reporting is not to compare results across countries but to inform system improvement within individual countries or country groups.

**Figure 2.9 A holistic framework to reporting**

| | UIS reporting | Tools for countries | Objective |
|---|---|---|---|
| 1 | **What content/ skills/abilities?** | Global Content Framework in reading and mathematics | To guide countries in the constructs they should include |
| | | Content alignment tool | To assess coverage |
| 2 | **How to implement?** Procedures and tools to report | Manual of good practices<br>Quick guides<br>Dashboard of learning assessments | To guide countries in implementation |
| | | Content alignment tool | To assess procedures |
| 3 | **Reporting**<br>1. What is the scale<br>2. What is the minimum proficiency level?<br>3. What options to link? | 1. Proficiency Framework definition for reading and mathematics<br>2. Definition of minimum proficiency<br>3. Toolkit for linking options<br>a. Test-based linking<br>b. Social moderation | To guide country reporting and implementation of relevant benchmarks |

*Source: UNESCO Institute for Statistics (UIS).*

# 3. Learning evidence for Indicator 4.1.1

SDG Target 4.1 covers the quality of primary and lower secondary education. The key concepts to measure include the quality of education and learning in two subject areas in early and late primary education and at the end of lower secondary education. The current global indicator for this target is the proportion of children and young people: i) in Grade 2 or 3; ii) at the end of primary education; and iii) at the end of lower secondary education who achieved at least a minimum proficiency level in (a) reading and (b) mathematics.

**Table 3.1 Summary of cross-national initiatives**

| Grade/age | Assessments |
|---|---|
| 1 | EGMA, EGRA |
| 2 | EGMA, EGRA, PASEC |
| 3 | EGMA, EGRA, LLECE |
| 4 | PILNA, LANA , PIRLS, TIMSS |
| 5 | SEA-PLM |
| 6 | LANA, PASEC, PILNA, SACMEQ, LLECE |
| 8 | TIMSS |
| 15-year-olds (Grade 7 or above) | PISA |
| 14- to 16-year-olds | PISA-D |
| 5- to 16-year-olds | ASER, Uwezo |

*Source: Treviño and Ordenes, 2017.*

The international initiatives that could help to inform SDG 4.1.1 are summarised in **Table 3.1** based on the target population/grade.

This chapter focuses on evidence from the major assessments presented in Table 3.1, as well as learning trends in a group of countries that are working with the Global Partnership for Education (GPE) to implement assessments needed to monitor and improve learning outcomes.

## 3.1 LEARNING EVIDENCE FOR INDICATOR 4.1.1 FROM REGIONAL ASSESSMENTS

The main regional assessments of the past decades to be analysed include:[17]

- Latin American Laboratory for Assessment of the Quality of Education (LLECE);
- CONFEMEN Programme for the Analysis of Education Systems (PASEC);
- SACMEQ (Southern and Eastern Africa Consortium for Monitoring Educational Quality); and
- Pacific Islands Literacy and Numeracy Assessment (PILNA).

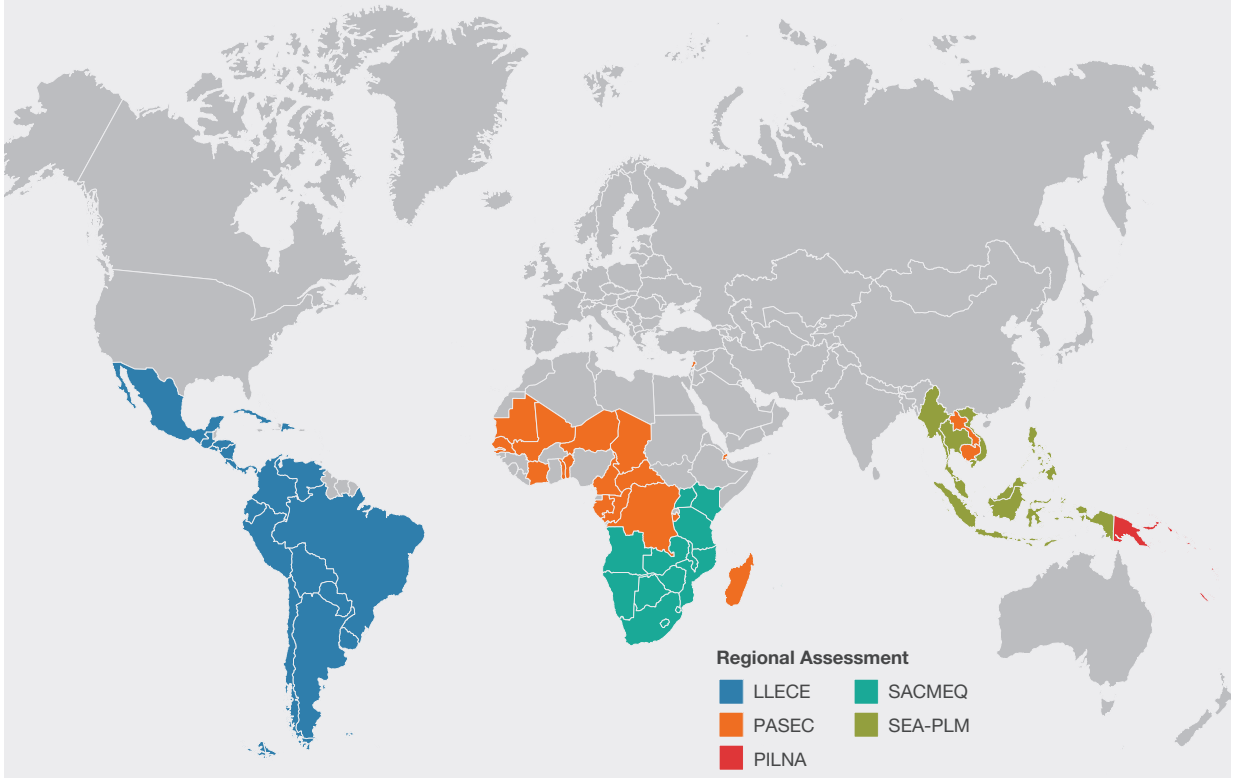
Country coverage is presented in **Figure 3.1**.

### 3.1.1 Latin American Laboratory for Assessment of the Quality of Education (LLECE)

LLECE is the leading quality assessment educational network in Latin America. It conducts the region's most representative evaluation of learning outcomes

17 One regional assessment is excluded: SEA-PLM (Southeast Asian Primary Learning Metric) because it will be first administered in 2019.

**Figure 3.1 The geographical coverage of regional assessments**




*Source: UNESCO Institute for Statistics (UIS).*

in primary education, the Regional Comparative and Explanatory Study (ERCE). This pan-Latin American network is made up of national-level directors of educational assessments in Latin American and Caribbean countries (LAC) and has its seat in the Regional Office for Education in Latin America and the Caribbean in Santiago, Chile. LLECE is also an important forum for analysing new approaches to educational quality and evaluation and for the discussion of learning outcomes. Importantly, it serves as an instrument for training and professional development of national technical teams.

Access to education is not the main challenge in LAC, where 95% of children are in school. However, ensuring that children learn well in the classroom and measuring their learning outcomes is critical to improving the quality of education in the region.

Recent research and surveys in preparation for SDG 4 have shown that, despite progress in the domain of access, the quality of learning is an issue in education systems, as well as the availability and access to educational resources.

Historically, the quality of assessments in most Latin-American countries has been uneven, with little knowledge of advanced student assessment. Many have government-conducted evaluations, but results have not been publicised. In fact, there has often been high resistance to publishing evaluation results and intense diplomatic efforts have been required to seek support from countries. The only country that has published the results of its assessments is Chile. In 1994, Chile's system was extended to all countries in LAC and established as a regional cooperation framework for the region. In 1994, LLECE had 15

founding members. Currently it includes 19 education systems in LAC.

### The work of LLECE in assessing the quality of education

The three main objectives of LLECE assessments are: to promote evidence-based education policy through the generation of (empirical) data on quality education and associated factors; to develop education assessment capacities; and to serve as a forum to generate and share ideas and best practices in education.

LLECE works through regional assessments within all contributing LAC countries to assess primary education in language, mathematics and science. So far, there have been three regional assessments: *Primer Estudio Regional Comparative y Explicativo* (First Regional Comparative and Explanatory Study) (PERCE), *Segundo Estudio Regional Comparative y Explicativo* (Second Regional Comparative and Explanatory Study) (SERCE) and *Tercer Estudio Regional Comparative y Explicativo* (Third Regional Comparative and Explanatory Study) (TERCE).

LLECE published PERCE on learning achievements in reading and mathematics among students in third and fourth grades in primary education. Its second regional study, SERCE, was implemented in 2006 and published in 2008. Among its innovations, SERCE applied the assessment of writing skills as well as a third discipline – science. LLECE's third study, TERCE, initiated in 2013, was a large-scale study of learning achievements implemented in 15 countries.[18] TERCE worked with its implementation partners, the Center for Measurement Mide UC, Pontifical Catholic University of Chile (UC), the Centre of Compared Policies of University Diego Portales (Chile) and the Colombian Institute for Educational Evaluation (ICFES), to develop the research tools and training that would lead to capacity building and the correct use of data.

The results of TERCE have been measured against those of SERCE (2006). This comparison demonstrated the changes which have occurred in the performance of the education systems of participating countries over the last seven years. Specifically, the results allowed learning achievements to be compared between pupils in Grades 3 and 6 in mathematics and reading tests. This was applied to all countries which participated in both studies. In addition, the natural sciences test results were compared for the eight countries for which data on both measurements were available. Participation in the science test was voluntary in SERCE and it was applied in only a few countries.

Other innovations in TERCE are the "national modules" of associated factors, which enabled countries to study in greater detail the factors which affect learning. It included a module to study the impact of the use of ICT on the quality of education and the relationship between nutrition and learning. TERCE also integrated the expertise of world-renowned experts in educational assessments.[19]

For the next study, ERCE 2019, there are two main innovations: a module to study in detail which pedagogical practices affect student learning and measure their impact; and a module to assess the development of socio-emotional skills, expanding the domains of evaluation of the LLECE's studies, according to the role of global citizenship education in the 2030 Agenda.

The national modules are very relevant for countries. For the implementation of ERCE-2019, one-half of the participating countries have developed a module to study specific topics of national interest. Some of these topics are: inter-culturality, perception of family support, pedagogical activities in science classes, impact of armed conflict, etc.

18 Argentina, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru and Uruguay, as well as the Mexican state of Nuevo Leon.

19 The databases for these three studies are available at: http://www.unesco.org/new/en/santiago/education/education-assessment-llece/perce-serce-databases/

### Reviewing and strengthening educational policies

In 2018 and 2019, the region will hold its fourth survey of regional assessment. A newcomer will be Bolivia. Cooperating with LLECE, Bolivia initiated an induction and capacity-building process for the Bolivian Ministry of Education and the Plurinational Observatory of Educational Quality (OPCE), strengthening the technical and institutional capacities of these entities. Within this framework, a national diagnostic was applied in order to serve as a record (baseline) for participation in ERCE 2019, which will be applied in 18 countries of the region. The ministry of education assumes responsibility for the enormous task of objectively evaluating what has been, is and will be

**Figure 3.2 Participation of Latin American countries in cross-national assessments**


*Source: LLECE.*

the Plurinational Educational System, and to rely on empirical evidence in the design of policies to improve the quality of education in LAC countries.

Bolivia's choice to participate in the LLECE evaluation model is a testament that the LLECE evaluation mode is better suited to the needs and characteristics of the region's countries; is based on national theoretical curricula and constructs; and makes it possible to address issues of specific relevance to Latin America. The study includes an evaluation of the factors associated with achievements in apprenticeships through questionnaires directed at school directors, teachers, families and students, with the objective of understanding how various socioeconomic and other factors affect the results. LLECE is acknowledged as creating a unique database which will help ministers of education in Latin America to make informed decisions on education policies based on the results of these investigations and the work of LLECE.

### 3.1.2 *Programme d'analyse des systems éducatifs de la CONFEMEN* (Programme of Analysis of Education Systems of CONFEMEN) (PASEC): The link between early school attendance and learning outcomes[20]

Over the past few decades, the world has focused on universal access to primary education. During the period, governments and the international community have been investing in school infrastructure, training teachers and developing learning materials. Children's attendance in school has increased, but many children are not learning (UIS, 2017g). In the new era of SDG 4, the focus is on learning quality and equity.

Many countries are promoting quality by improving the monitoring of learning by national, regional and international learning assessments, and by developing targeted programmes that improve teaching and learning. The first part of this section will introduce the work of CONFEMEN through its regional assessment, PASEC, to monitor Indicator 4.1.1. The second part will look at the factors that improve learning outcomes, more specifically the link between early school attendance and learning outcomes. The data come from the PASEC 2014 assessment.

#### The CONFEMEN Analysis Programme of Education Systems

PASEC was created in 1991 to conduct assessments of education achievement at the primary school level with a focus on basic education.[21] From 1991 to 2012, PASEC carried out national evaluations in nearly all the francophone countries of sub-Saharan Africa, in Lebanon and in three South Asian countries (Cambodia, PDR and Viet Nam).

Starting in 2012, CONFEMEN reformed PASEC to direct its methodology towards creating international assessments, grouping together several countries by means of standardised surveys, allowing for participating countries to make comparisons amongst each other. With these new standards, PASEC undertook its first international evaluation, called PASEC 2014. Ten countries participated in this evaluation: Benin, Burkina Faso, Burundi, Cameroon, Chad, Congo, Côte d'Ivoire, Niger, Senegal and Togo. An international report which analysed data from the countries and gave comparative results was developed and published. National reports were also published. These national reports presented the results of the assessments at the national level, on the one hand through the comparison of the country to the other education systems, and on the other hand through a comparative analysis of the various school regions of the education systems. In these national reports, results were analysed in each of these contexts.

According to PASEC 2014 results, more than 70% of students at the beginning of primary education

20 Written by Hilaire Hounkpodoté, PASEC Coordinator, CONFEMEN.

21 The *Conférence des ministres de l'éducation des états et gouvernements de la Francophonie* (CONFEMEN), created in 1960 and including 44 countries and governments, supports, according to its mission, member countries in the improvement of the quality of their education systems through CONFEMEN's PASEC.

(Grade 2) have not attained the expected skill level in reading/writing, and more than 50% have not attained the expected skill level in mathematics. At the end of primary education (Grade 6), nearly 60% of students are below the expected skill levels in both subjects. These results demonstrate that, despite progress in expanding access to schools, the quality of learning is an issue. Furthermore, the results also show that in many schools the availability of and access to educational resources are a challenge.

A good follow-up according to SDG 4 suggests that countries assemble and arrange data for targeted indicators. National, regional and international assessments are therefore essential. After the first evaluation in 2014, PASEC started to prepare for its second round of international assessments, PASEC 2019, which will include 15 countries: Benin, Burkina Faso, Burundi, Cameroon, Chad, Congo, Côte d'Ivoire, Democratic Republic of the Congo, Gabon, Guinea, Madagascar, Mali, Niger, Senegal and Togo. The testing of instruments and procedures for this second evaluation took place in April, May and June of 2018. The final data collection is planned for April and May of 2019.

PASEC assessments are based on the measurement of skills in reading and mathematics at the start and end of primary education (Grades 2 and 6) and the analysis of factors that contribute to academic success in order to propose ideas and actions for improvements. PASEC also reinforces the capacities of countries. Since its creation, PASEC has strengthened the capacities of national teams concerning several themes, including the creation of instruments, sampling, data processing and development of reports.

### Academic achievement factors: The link between attending preschool and primary school learning results

Data from PASEC 2014 showed that household inequality, school pathways and school and class characteristics, particularly educational resources, might lead to differences in success at the beginning and end of schooling. The summary of the data demonstrates that a number of factors favour school success. For instance, attending urban schools, the availability of educational resources, attending preschool, not repeating a grade, literacy of one parent and the availability of necessary educational resources in schools and classes are just some of the factors that are conducive to academic achievement.

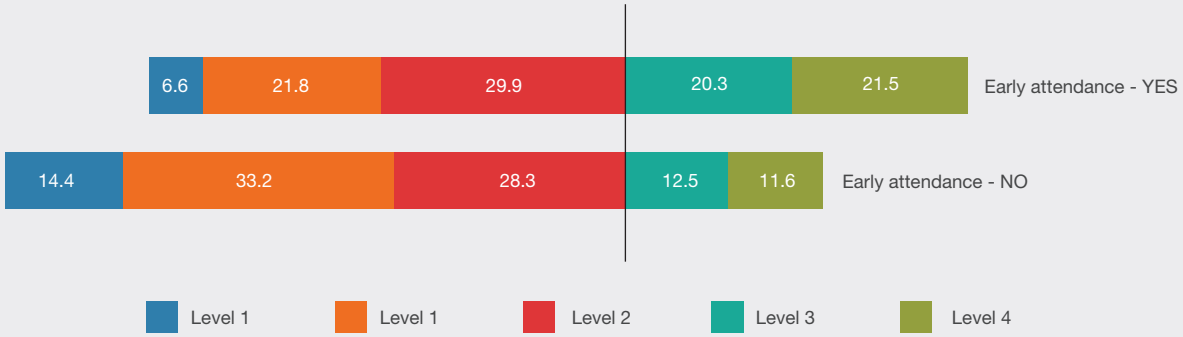
Among the factors mentioned above, attending preschool is being discussed more and more within education systems. This topic of providing preschool education to children was at the heart of the dialogue during the 58th CONFEMEN conference held in May 2018 in New Brunswick, Canada. Early childhood years, the period between birth and six years old, is understood today as a crucial period for the development of young children, both from the point of view of physical health and that of motor, socio-emotional, cognitive and language development. In this respect, preschool education prepares children to approach their first instances of learning in good condition. This preparation is even more important for children coming from underprivileged backgrounds.

As the PASEC 2014 international report highlights, the idea of preschool education is very different from one country to another. Teaching programmes, the type of teaching and even the language of instruction can vary. According to PASEC 2014 data for the 10 participating countries *(see **Figure 3.3**)*, between 10% and 50% of enrolled students in Grade 2 attended preschool before starting primary school. The percentage of students who attended preschool and who have attained the expected skill level in reading/writing is 41.8%, compared to 24.1% for those who did not attend preschool.

In Grade 6, between 12% and 46% of students attended preschool, with a relatively high percentage in Cameroon, Benin, Senegal and Congo.

The gross analysis of performance differences between students who did attend preschool and

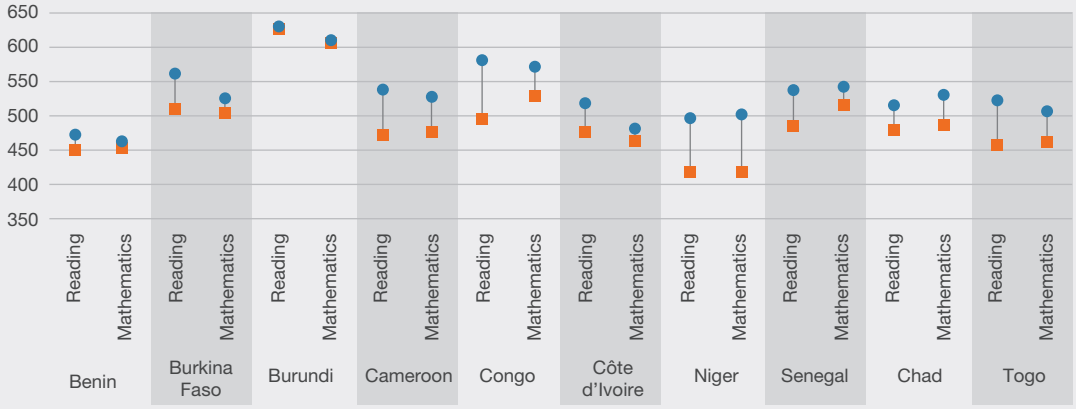
**Figure 3.3 Percentage of students attending or not attending preschool and their corresponding skills levels at the start of primary school for reading and writing**



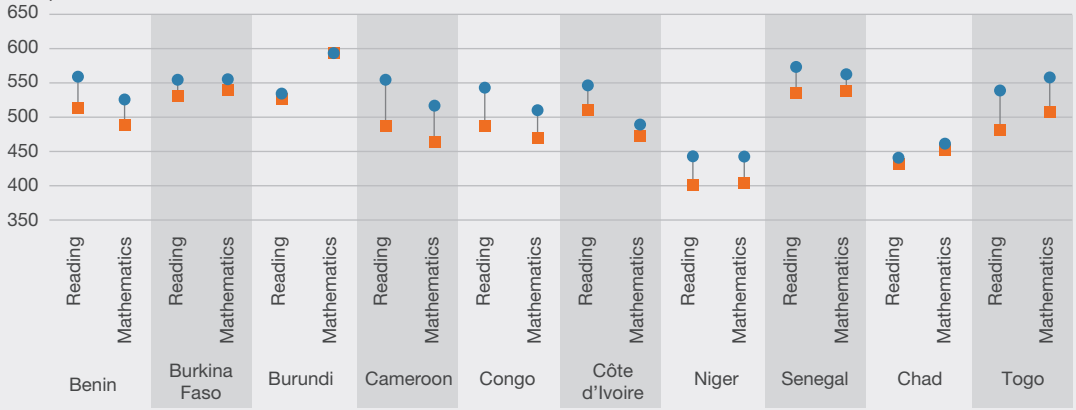

*Source: PASEC.*

**Figure 3.4 Gross difference between students who attended preschool and those who did not**

Performance of students in reading and mathematics at the start of school depending on their attendance in kindergarten or pre-school




Performance of students in reading and mathematics at the end of school depending on their attendance in kindergarten or pre-school

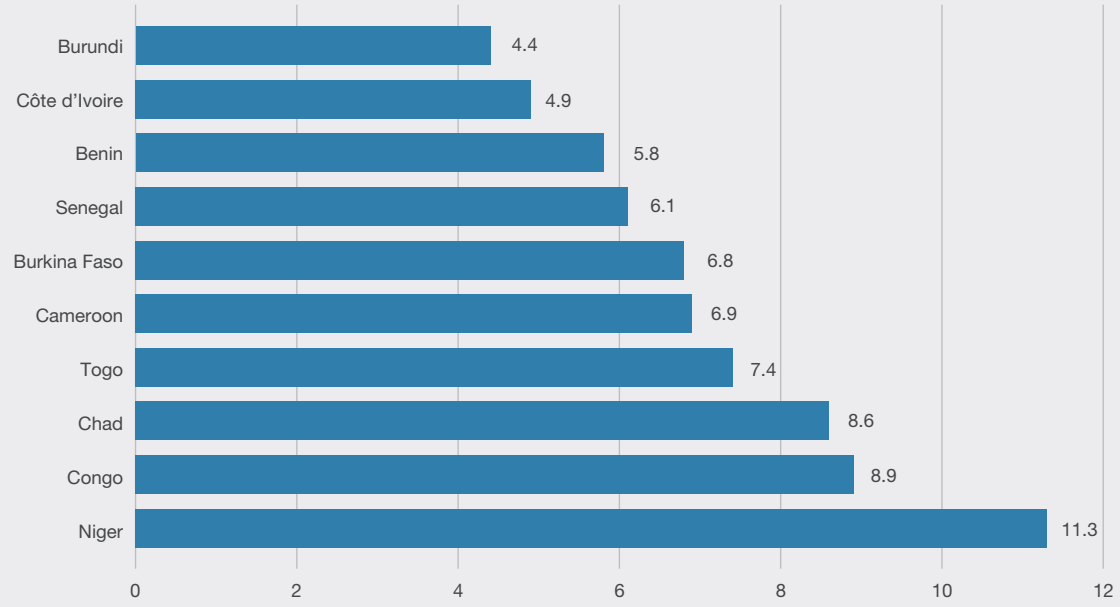



*Source: PASEC, 2014.*

Mean score for students attending Kindergarten or pre-school

Mean score for students not attending Kindergarten or pre-school

**Figure 3.5 Average socioeconomic level gap between students who attended preschool and those who did not**




*Source: PASEC, 2014.*

those who did not demonstrates that in most countries students who did attend preschool perform better than those who never went *(see **Figure 3.4**)*.

A more elaborate analysis of the data shows that students who attended preschool have on average a higher socioeconomic status than students who did not attend preschool. In addition, students who attended preschool predominantly live in urban areas.

The difference between the socioeconomic level of students who attended preschool and those who did not is shown in **Figure 3.5**. As seen in the figure, the greatest difference in socioeconomic level between the two types of students is 11.3 in Niger. This difference is 8.9 in Congo and 8.6 in Chad. In other countries, the gap varies between 4.4 and 7.4.

The analysis of the link between preschool attendance, controlling for socioeconomic levels of students' families, schools in urban areas and student performance shows that in most countries the link remains positive and significant between preschool attendance and student performance. While the gap between performances of the two types of students decreases when controlling for the variables mentioned above, the gap is still relatively significant *(see Figure 3.4).*

At the start of primary schooling, the gap remains significant for reading/writing in Cameroon (74.9), Senegal (47.3) and Togo (40.6). In Benin, the gap is 20 points, while in the Congo the gap is 37.4 points. The exceptions are Burkina Faso and Burundi, where the link is not significant. In mathematics, the gap remains important in Cameroon and Niger but is not significant in the six other countries (Benin, Burkina Faso, Burundi, Congo, Côte d'Ivoire and Senegal).

At the end of primary schooling, the largest gaps in reading/writing are in Cameroon (39.9), Togo (33.0), Congo (25.3), Benin (24.6), Senegal (18.1) and Côte d'Ivoire (17.1). In mathematics, the gaps are only

found in Cameroon (29.8), Togo (26.2), Benin (24.6), Congo (21.3) and Niger (22.7).

Analysing the link between attending preschool and students' academic performance allows us to ask questions and examine different lines of thought. These include:

- Integration of preschool in national curricula. Should preschool education be included in the basic education cycle/primary education?
- Coordination between the preschool education programme and those of primary or basic education.
- Harmonisation of management approaches and learning content in a context of diverse preschool opportunities.
- Qualifications of preschool management staff.

The situation regarding the quality of education systems in sub-Saharan Africa concerns all actors in the field of education. The culture of learning assessments and the inclusion of the results in education politics and decisionmaking are integral for countries if they want to promote inclusive and quality education for all by 2030.

### 3.1.3 Southern and Eastern African Consortium on Monitoring Educational Quality (SACMEQ)

The Southern and Eastern African Consortium on Monitoring Educational Quality (SACMEQ) is a collaborative network of 15 ministries of education who conduct standardised assessments to measure the quality of education in countries and jurisdictions of Southern and Eastern Africa. SACMEQ research and assessments are informed by policy concerns identified by ministers from member countries. The Consortium's mission is to develop the capacities of education planners to monitor and evaluate the conditions of schooling and the quality of their basic education systems, and to generate research-based information that can be used by decisionmakers to plan for improvements in the quality of education.

SACMEQ has conducted four nationally-representative, school-based surveys in member countries. The surveys, SACMEQ I (1996), SACMEQ II (2000), SACMEQ III (2007) and SACMEQ IV (2013), test learners and teachers in numeracy and literacy in Grade 6 and collect extensive background information on the schools and home environments of students. Assessments and background questionnaires provide information on school characteristics (e.g. location, enrolment, resources, principal's qualifications), learning characteristics (e.g. age, gender, attendance, nutrition, socioeconomic status) and teacher characteristics (e.g. age, gender, qualifications, behaviour, in-service training).

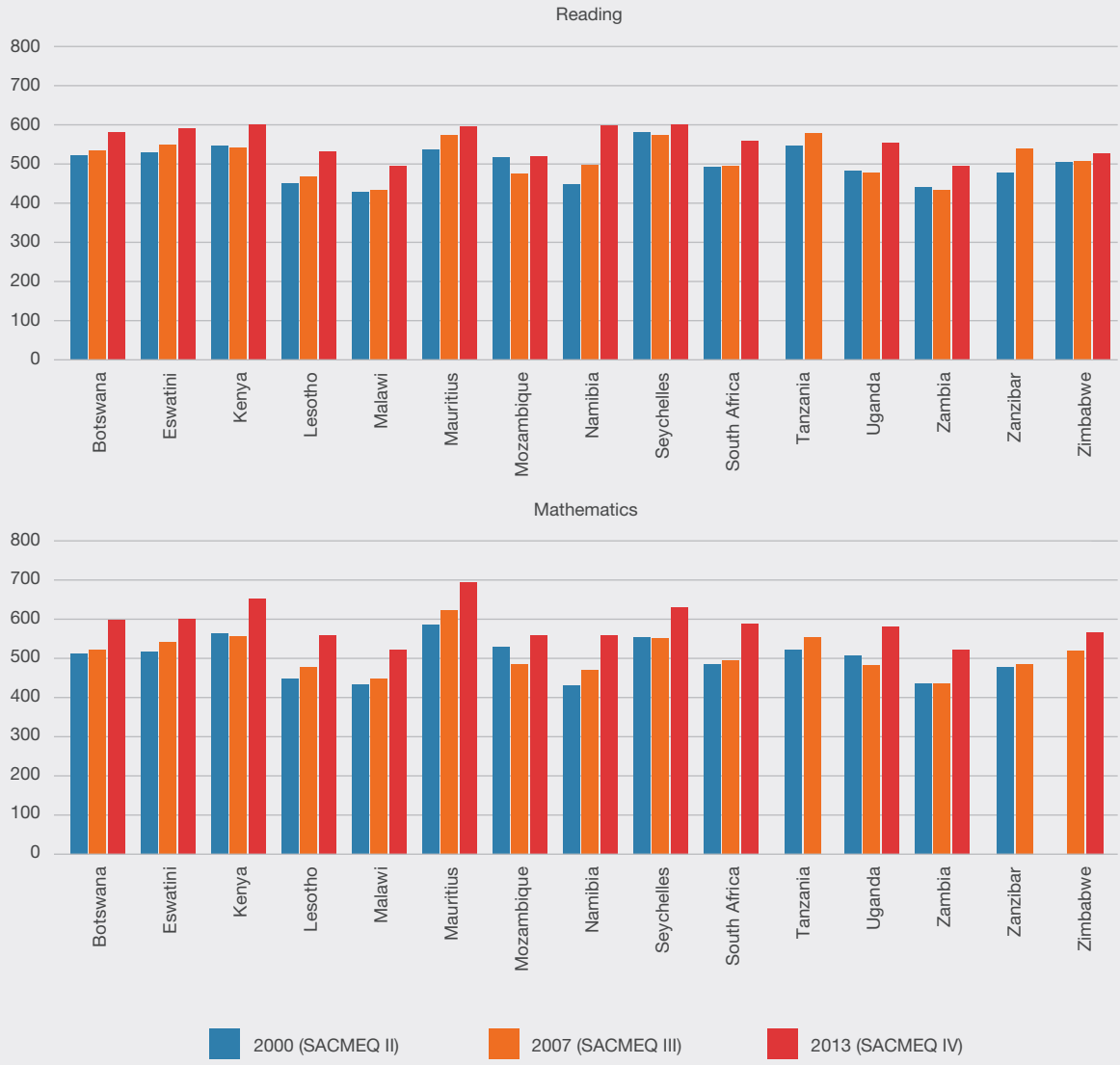
SACMEQ reading and mathematics test frameworks cover curriculum topics that are common across member countries. Testing instruments are developed to examine changes in the performance of a single education system across several points in time and to explore the differences in the performance of education systems at a single point. Therefore, SACMEQ assessments are comparative cross-nationally and within countries over time. Samples and sample sizes are designed to ensure that estimates are reported with 95% levels of confidence and are nationally-representative. Concerning test construction and scoring, item response theory (IRT) models facilitate the generation of valid comparisons of reading and mathematics achievements across and within SACMEQ countries.

The results described in **Figure 3.6** show improvement in all countries with respect to the initial year, although there is no report describing the technical features, including longitudinal linking, that are common to various regional assessments. To date no consolidated report has been published.

### 3.1.4 Pacific Islands Literacy and Numeracy Assessment (PILNA)

The Pacific Islands Literacy and Numeracy Assessment (PILNA) is a regional assessment of literacy and numeracy administered at the end of

**Figure 3.6 Mean achievement scores by country (SACMEQ II-IV), Grade 6**




*Source: UNESCO Institute for Statistics (UIS).*

four and six years of formal education. PILNA is administered every three years, in 10 languages across 15 participating countries, to over 40,000 students in more than 700 schools. PILNA was created in 2012 as a one-time measure of literacy and numeracy in response to concerns from education ministers that students in the Pacific were not performing well. The results of PILNA 2012 validated the ministers' concerns, leading them to ask for a follow-up assessment to determine the impact of new literacy and numeracy interventions.

In 2015, the New Zealand Ministry of Foreign Affairs and Trade provided generous funding support to enable the Educational Quality and Assessment Programme (EQAP) of the Pacific Community (SPC) to develop the current PILNA initiative, with Australia joining the funding initiative in 2018. Beginning in

2015, PILNA adopted a collaborative governance structure and strong technical partnership with the Australian Council for Educational Research (ACER) to ensure the development of a robust large-scale assessment with high-quality instruments and valid and reliable results that are useful to participating countries.

A steering committee guides the administration of PILNA, which consists of a chief executive officer from each country's ministry of education, representatives from the New Zealand and Australian governments and the director of EQAP. The steering committee has represented the strategic priorities of participating countries and has engaged in high-level discussions on behalf of the ministries. The committee has decided how to communicate and use the PILNA results nationally and regionally and endorsed a data-sharing commitment, which outlined how data should be employed so as not to inform league tables or comparisons. In addition, committee members have worked to ensure the accessibility of results, including presenting them in a meaningful manner to governments, schools and teachers by SPC officers and disaggregating data by gender, school authority and locality in national reports. Overall, the steering committee has engaged with countries at every stage of the PILNA process, including item development, field trials, data analysis and reporting.

## 3.2 LEARNING EVIDENCE FROM GLOBAL ASSESSMENTS

This section analyses tests administered by two global organizations: the IEA and the Organisation for Economic Co-operation and Development (OECD).

### 3.2.1 Measuring SDGs and improving education with IEA studies[22]

The IEA is a non-profit, international, scientific society that conducts pedagogical research worldwide. More than 60 countries are represented in its membership, and over 100 education systems participate in IEA studies.

IEA studies are designed by educators for educators to answer questions such as: What do students know and what can they do? Is student achievement improving over time? What practices and policies are associated with student achievement? The aim is to inform and help all educators support upcoming generations become more successful learners rather than fuel a competition among education systems.

All of IEA's assessments are grade-based and curriculum-rooted. There are two major populations of IEA research interest: Grade 4 (10-year-old students at the primary level) and Grade 8 (14-year-old-students at the lower secondary level). Some IEA studies target additional grades and groups.

IEA studies consider the processes and outcomes of education and draw upon the notion of "opportunity to learn" in order to understand the linkages between the intended curriculum (what policy requires), the implemented curriculum (what is taught in schools) and the achieved curriculum (what students learn). This model is expressed in the frameworks that precede the instrument development and data collection in all studies (Mullis and Martin, eds., 2014, 2015 and 2017 and Schulz et al., 2016). IEA studies measure student achievement in subjects such as mathematics and science (TIMSS), reading (PIRLS), civic and citizenship education (ICCS), and computer and information literacy (ICILS). The cyclical design of these studies enables the measurement of trends in educational achievement across multiple contexts. All study results and data are open access and freely available on the IEA website.[23]

IEA studies measure student achievement by administering tests to a sample of students who have been selected as representative of national populations at a specific grade. This is critical so that the sample represents the whole target populations

22 Written by Paulína Koršnáková, Senior Research and Liaison Adviser, and Dirk Hastedt, Executive Director, IEA.

23 www.iea.nl/data

**Table 3.2 Overview of IEA studies and the SDG targets they can support**

| Study | Target grade(s) | Instruments and respondents | Study years | SDG targets |
|---|---|---|---|---|
| TIMSS | 4 (10-year-olds)<br>8 (14-year-olds) | Test and questionnaires for students; parental questionnaires (Grade 4 only); questionnaires for school principals and teachers; national context questionnaires. | 1995, 1999, 2003, 2007, 2011, 2015 | 4.a, 4.c<br>4.1<br>4.2<br>4.4<br>4.5 |
| PIRLS | 4 (10-year-olds) | Test and questionnaires for students; parental questionnaires; questionnaires for school principals and teachers; national context questionnaires. | 2001, 2006, 2011, 2016 | 4.a, 4.c<br>4.1<br>4.2<br>4.5 |
| ICILS | 8 (14-year-olds) | Test and questionnaires for students; questionnaires for school principals, information and computer technology coordinators and teachers; national context questionnaires. | 2013, 2018 | 4.a, 4.c<br>4.4 |
| ICCS | 8 (14-year-olds) | Test and questionnaires for students; regional modules; questionnaires for school principals and teachers; national context questionnaires. | 2009, 2016 | 4.a, 4.c<br>4.7 |

*Source: IEA.*

within the participating education systems and may serve as a mirror for education policy, practices and outcomes.

Background information is collected about factors that affect learning, including school resources, student attitudes, instructional practices and support at home. Information is also collected from school principals, teachers, students and, in some studies, parents and policymakers. The resulting data are organized and stored in an international database, and the datasets linked to specific studies are well described in user guides (Foy, ed., 2018).

National research coordinators ensure that test instruments and procedures are appropriate for their students and fit the context of their country. Assessment questions are pre-tested (referred to as pilot and field testing), and issues are addressed before the main assessment is administered. The IEA also makes every effort to safeguard the quality and comparability of data through careful planning and documentation, cooperation among participating countries, standardised procedures and rigorous attention to quality control (Martin et al., 2017).

### IEA studies and the SDGs

IEA's open access datasets are recognised by UNESCO as a solid evidence base for researchers, educators and policymakers interested in monitoring progress toward the SDGs *(see **Table 3.2**)*.

In the following sections, we discuss some of the key results from IEA studies and their relevance for SDG monitoring.

### TIMSS and PIRLS

Two decades of TIMSS results (1995-2015) reveal important trends, several of which are noted here. First, more countries have registered increases rather than decreases in average student achievement scores in Grade 4 and 8 mathematics and science (Mullis, Martin and Loveless, 2016). More students are also now reaching the most challenging benchmarks,

**Table 3.3 Percentage of children and young people at the end of primary (Grade 4) and the lower end of secondary school (Grade 8) who achieved at least a minimum proficiency level, equivalent to the low achievement level in TIMSS and PIRLS, in mathematics and reading**

| | Primary education (Grade 4) | | Lower secondary education (Grade 8) |
|---|---|---|---|
| | Mathematics (TIMSS 2015) | Reading (PIRLS 2016) | Mathematics (TIMSS 2015) |
| Percentage of students who achieve at least a minimum proficiency level | 93 | 96 | 84 |
| Number of countries included | 49 | 50 | 39 |

**Notes:** Based on the World Bank list of economies from June 2018, 38 out of the 49 participating countries in TIMSS 2015 Grade 4 were classed as high-income countries, 8 were upper-middle-income countries and only 3 were lower-middle-income countries (Georgia, Indonesia and Morocco). No low-income countries participated in TIMSS 2015.
For Grade 8 participants of TIMSS 2015, two-thirds were classified as high-income countries (26) and one-third as middle-income countries (13). Among the 50 participating entities in PIRLS 2016, there was 42 high-income countries.
*Source: Adapted from Mullis et al., 2016b and 2017b and Martin et al., 2016.*

and gender gaps in student achievement are decreasing. These overall improvements in educational achievement trends are accompanied by additional gains, such as improved school environments (e.g. safer schools), better educated teachers, more support for teachers' professional development and better curriculum coverage (Mullis, Martin and Loveless, 2016).

The 15-year trends in PIRLS results (2001-2016) also show more increases than decreases in student achievement. Internationally, there are more proficient readers than there were 15 years ago (Mullis et al., 2017b). The gender gap in Grade 4 reading achievement has favoured girls since 2001 and does not appear to be closing. However, there are some examples of educational systems with no significant gender differences on overall PIRLS scores. These include Portugal and Macao Special Administrative Region of China in the PIRLS 2016 assessment and Denmark, Italy and Portugal in an innovative online reading assessment, ePIRLS 2016.

Second, TIMSS and PIRLS results are useful tools for monitoring progress towards inclusive, equitable and quality education as measured by the SDGs.

### Indicator 4.1.1

**Table 3.3** summarises TIMSS 2015 and PIRLS 2016 results relating to Indicator 4.1.1: the proportion of children (Grade 4) and young people (Grade 8) achieving at least a minimum proficiency level in reading and mathematics.

The TIMSS and PIRLS low international benchmarks represent basic functions and competencies and have been identified as the most appropriate to measure "SDG minimum proficiency level". For example, students who meet this level in TIMSS Grade 4 mathematics (average of 93% across all countries) can add and subtract whole numbers, have some understanding of multiplication by one-digit numbers, can solve simple word problems and have some knowledge of simple fractions, geometric shapes and measurements. Meanwhile, 95% of Grade 4 students reached the TIMSS 2015 minimum proficiency level in science by demonstrating that they have some basic knowledge of the interaction of living things with their environments and its application related to human health (Martin et al., 2016).

Students who achieve the low international benchmark for PIRLS can locate, retrieve and reproduce explicitly-stated information from a text, make straightforward inferences and begin to interpret story events and central ideas. The 96% of students who achieved at least a minimum reading proficiency level in PIRLS 2016 is an average from all 50 participating countries *(see Table 3.3).* While this is a positive result, there is still much work to do in improving literacy levels for countries at the lower end of the scale. The situation is particularly challenging in three countries (Morocco, Egypt and South Africa) where considerably fewer than 50% of Grade 4 students achieved the minimum reading proficiency level in PIRLS 2016 *(see **Figure 3.7**).*

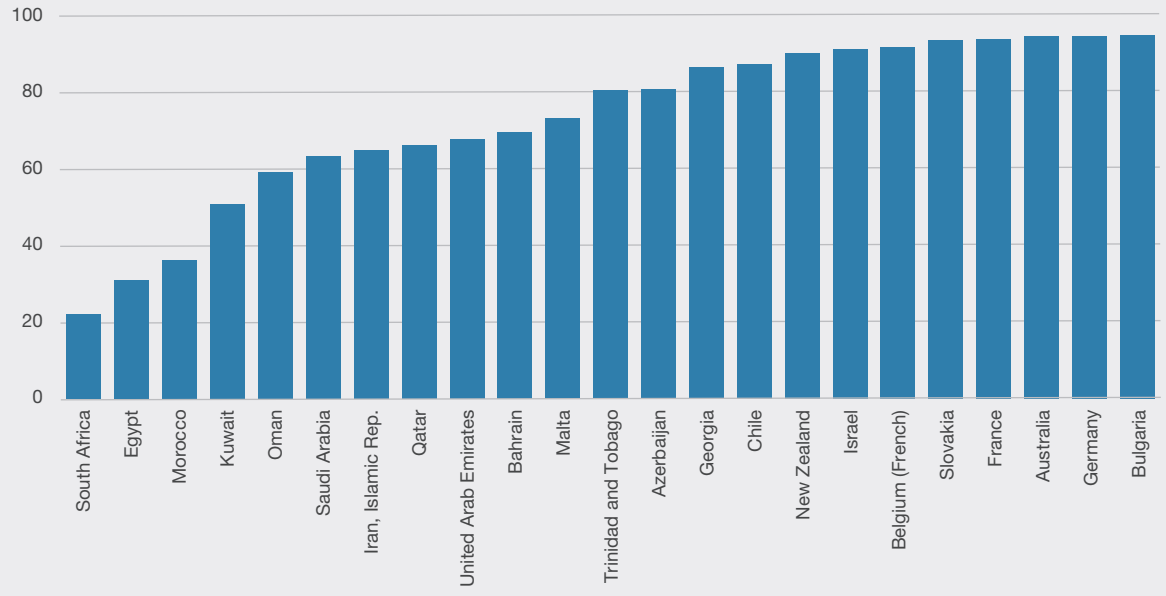
### SDG Targets 4.a and 4.c

In TIMSS 2015, the vast majority of Grade 4 students (90%) were in safe school environments (SDG Target 4.a) according to their principals and teachers (Mullis et al., 2016). The 10% of students who attended schools with disorderly environments had much lower mathematics and science achievements than their counterparts in safer schools – 468 vs. 512 scale score points – or close to one-half of a standard deviation.

It is worrying, however, that 45% students in TIMSS 2015 reported that they were bullied monthly or weekly. The results of the secondary analysis performed by Rutkowski and Rutkowski (2018) showed that bullying is not isolated to one country but is an international phenomenon that tends to have an impact on the mathematics achievement of those students who are bullied.

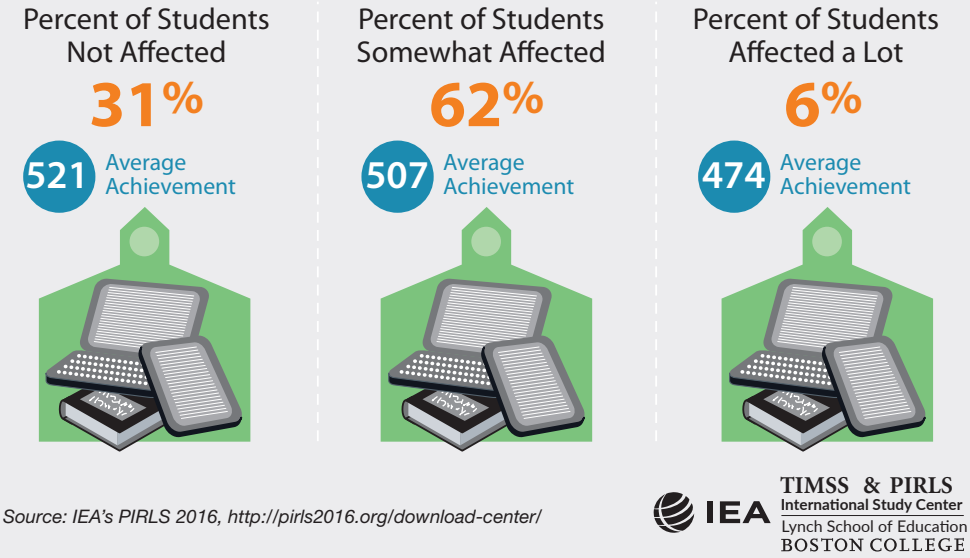
PIRLS context questionnaires help to monitor problems with school conditions and resources (SDG Target 4.c). Based on principals' reports, only 31% of students who participated in PIRLS 2016 were not affected by any reading resource shortages,

**Figure 3.7 Percentage of Grade 4 students who performed at or above the minimum reading proficiency level (400 scale score points) in PIRLS 2016**




**Note:** The figure shows the 23 countries where less than 95% of students performed at or above the minimum reading proficiency level.
*Source: IEA Research and Analysis Unit, based on PIRLS 2016 data.*

**Figure 3.8 Instruction affected by reading resource shortages according to principals' reports, PIRLS 2016**




*Source: IEA's PIRLS 2016, http://pirls2016.org/download-center/*

IEA TIMSS & PIRLS International Study Center Lynch School of Education BOSTON COLLEGE

and they achieved an average score of 521 points. In contrast, the 6% of students who were strongly affected by shortages in reading resources scored significantly lower, at 474 points, which was below the international average *(see **Figure 3.8**)*. This is about one half of a standard deviation.

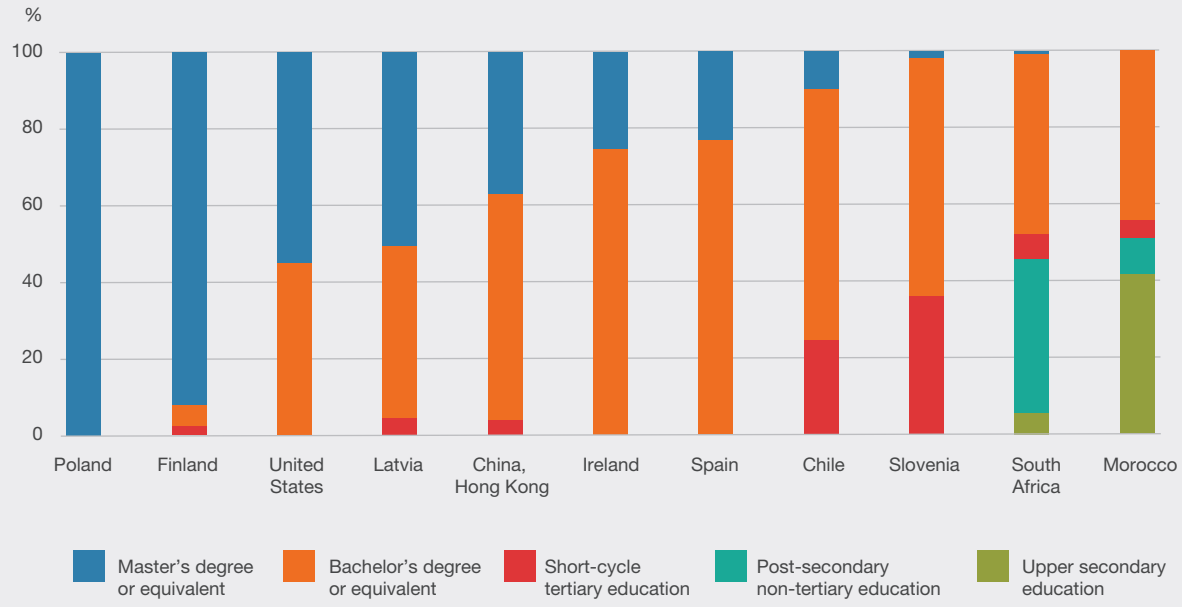
UNESCO/IEA (2017) provides a comprehensive overview of the scope and depth of information available which can be used to improve teaching and learning. For example, while there is no international consensus on the definition of a qualified teacher, a teacher's highest level of formal education can serve as one indicator. **Figure 3.9** demonstrates some of the variation in the qualification levels of reading teachers (Indicator 4.c.1) for Grade 4 students who participated in PIRLS 2016. These data give useful insight for countries to monitor the proportion of their teachers who have received the training required for instructing a given level in their country.

### SDG Targets 4.2 and 4.5

Children with access to pre-primary education (Indicator 4.2.2) tend to show higher achievement in schools, including in PIRLS results. Monitoring PIRLS results over time demonstrates that, in many countries, attendance rates in pre-primary schools have increased *(see **Figure 3.10**)*. This is an encouraging indication of progress towards achieving SDG Target 4.2.

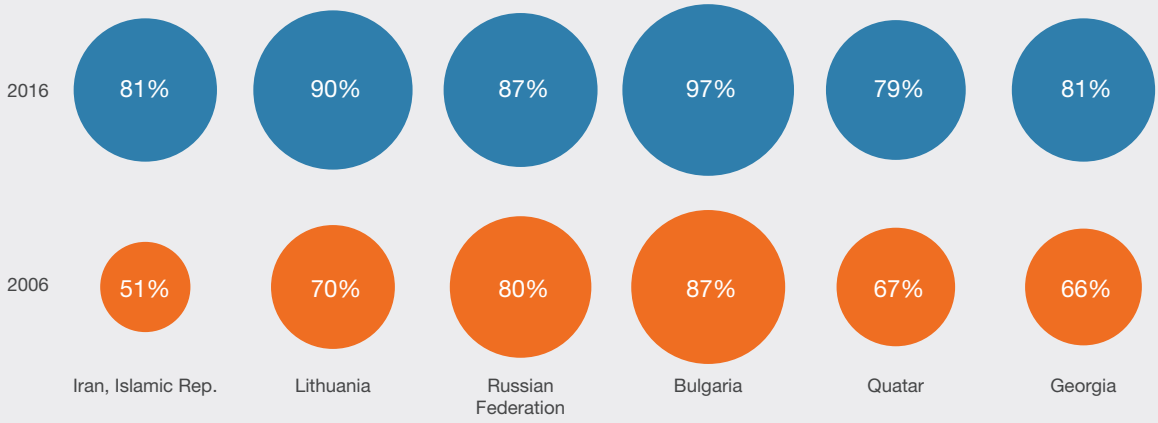
PIRLS 2016 data reaffirmed gender disparities in educational achievement (SDG Target 4.5) in favour of girls. In contrast, a UNESCO publication, *Cracking the Code* (UNESCO, 2017a), looked at a smaller sub-set of trend data based on 17 countries that participated in TIMSS between1995 and 2015. The analysis focused on the factors that influence women's under-representation in the science, technology, engineering and mathematics (STEM) professions. Based on this sub-sample of TIMSS results, the authors reported a slight improvement in reducing gender differences between average TIMSS scores for girls and boys.

**Figure 3.9 Percentage of Grade 4 students in selected PIRLS 2016 countries who were taught by teachers of different qualification levels**




*Source: UNESCO/IEA, 2017.*

**Figure 3.10 Percentage of pre-primary school attendance in children who participated in PIRLS 2006 and PIRLS 2016**




*Source: UNESCO/IEA, 2017.*

### ICILS and Indicator 4.4.1

As the first study to create an international benchmark of digital literacy proficiency levels and to investigate the factors that influence these skills in young people (Fraillon et al., 2013), ICILS offers insights into Indicator 4.4.1: the proportion of youth with ICT skills.

Results from ICILS 2013 showed that only 2% of students displayed an application of critical thinking when searching for information online (Level 4 of the digital proficiency scale, see **Figure 3.11**). This result highlights that children who belong to a "digital native" generation still need to be taught how to interact with digital information. Encouragingly, 84% of students reached Level 1 on the digital proficiency scale, indicating that they can master basic software commands to access files and complete routine text and layout editing tasks (Fraillon et al., 2014).

### ICCS and SDG Target 4.7

Since 1970, the IEA has investigated the ways in which young people (Grade 8 students) are prepared to undertake their roles as global, democratic citizens. This effort led to the ICCS measuring students'

**Figure 3.11 Percentage of students in ICILS 2013 who reached specific proficiency levels of digital literacy (averages across 21 participating education systems)**


**Note:** Averages of 21 participating education systems.
*Source: ICILS, 2013, https://www.iea.nl/sites/default/files/studies/ICILS_2013_infographic.pdf*

understanding of civic systems and principles, their beliefs, attitudes and behaviours. The study aims to combat low levels of tolerance and to encourage student participation and engagement.

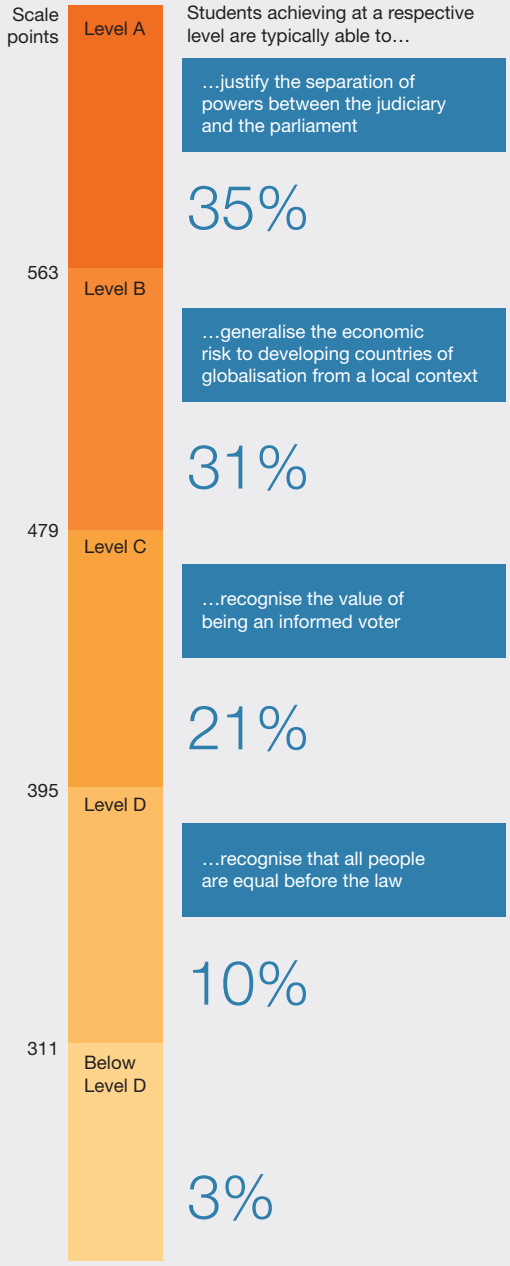
By addressing cognitive- and affective-behavioural constructs related to civic and citizenship education, ICCS inspired the development of the SDG thematic Indicator 4.7.4: the percentage of students showing adequate understanding of issues relating to global citizenship and sustainability.

Based on the measurement strategy for SDG Target 4.7, global citizenship education (GCED) is tentatively defined as "any educational effort that aims to encourage the acquisition of skills, values, attitudes and behaviours to empower learners to assume active roles to face and resolve global challenges and to become proactive contributors to a more peaceful, tolerant, inclusive and secure world" (UIS, 2017h, p.3).

There is an ongoing discussion about what should be classed as the minimum proficiency level of GCED. Here, we suggest that Level C in the ICCS civic knowledge scale should be considered. Students who reach Level C in their civic awareness "typically demonstrate awareness of citizens' capacity to exert influence in their own local context". In addition, they already possess the capacity of the previous proficiency Level D, so they also "recognise examples of respect for the rights of others, and they may see these rights as motivation for citizenship engagement" (Schultz et al., 2018).

**Figure 3.12** illustrates students' proficiency levels and the average percentage of students who achieved each level across the 21 countries that participated in ICCS 2016. In this cycle, 13% of students did not reach Level C. It will be interesting to monitor whether these percentages improve in the next cycle of ICCS in 2022.

**Figure 3.12 Percentage of students who achieved each proficiency level in ICCS 2016**




*Source: ICCS, 2016. https://iccs.iea.nl/cycles/2016/findings/single-finding/news/iccs-2016-infographics-civic-knowledge-levels-and-trends/*

### Using TIMSS Science for Indicator 4.7.5

Indicator 4.7.5 relates to the "percentage of 15-year-old students showing proficiency in knowledge of environmental science and geoscience", areas that are partly covered by the TIMSS Grade 8 science framework.

TIMSS has been measuring trends in mathematics and science achievement at Grades 4 and 8 (and partly also the final grade of secondary education) since 1995. TIMSS assessments use the curricula of participating countries as a basis to investigate how countries are providing educational opportunities in mathematics and science to their students. Additionally, TIMSS investigates the factors related to how students are using these opportunities.

Currently 40 countries and 5 benchmarking entities from all over the world are participating in the TIMSS 2019 Grade 8 assessment. A similar number of countries participated in previous cycles of the assessment. Hence, mathematics and science achievement scales and international proficiency levels are well established and widely recognised.

The current cycle of TIMSS is focusing on converting to a digital format, allowing inclusion of additional practical tasks and experiments, such as a plant growth experiment, which can be used to more thoroughly assess student knowledge in the curriculum areas covered by the TIMSS frameworks. The TIMSS science framework in Grade 8 includes the content dimensions of biology, chemistry, physics and earth science, covering a globally-relevant perspective as the assessment framework is based on the national curricula of participating countries. The science part of the TIMSS Grade 8 main assessments typically consist of about 225 items, with only a fraction administered to each of the students to avoid overburdening. Currently, 338 new (paper) science items are in field trials to test their suitability to replace the released item blocks in the 2019 main data collection.

For Indicator 4.7.5, the content domains of biology and earth science are regarded as especially relevant. For all content sub-domains, separate scale scores are calculated. Each of the content areas include several major topics that are described by specific objectives. Objectives represent typical performances expected of the students and are assessed in three different cognitive domains (knowing, applying and reasoning).

In biology, two out of the six topic areas covered by the TIMSS science framework ("ecosystems" and "human health") are particularly relevant for Indicator 4.7.5. Students are assessed in terms of their understanding related to processes and interactions in ecosystems, topics that are seen as an essential basis for thinking about how to develop solutions to many environmental challenges. Furthermore, students should get a "science-based" understanding of human health" in order to improve the conditions of their lives and the lives of others" (p. 40). A more detailed description of the framework for the above-mentioned two topics can be found in Mullis and Martin (2017a).

In earth science, out of the four topic areas covered in the TIMSS framework, items related to the topics "earth's resources, their use and conservation" will be specifically relevant to the measurement of Indicator 4.7.5. The objective here is that "students should demonstrate knowledge of earth's resources and their use and conservation, and relate this knowledge to practical solutions to resource management issues" (Mullis and Martin, 2017a).

### Conclusions

The results reported here are just a few examples of the important insights into student achievement that IEA studies provide and how they relate to progress towards the SDGs. The studies continue to evolve to keep pace with the dynamic development and complexity of the education systems that they monitor. For example, upcoming cycles of ICCS will include measures of global citizenship

**Figure 3.13 PISA cycles**

| 2000 | 2003 | 2006 | 2009 | 2012 | 2015 | 2018 |
|---|---|---|---|---|---|---|
| **Reading** | Reading | Reading | **Reading** | Reading | Reading | **Reading** |
| Mathematics | **Mathematics** | Mathematics | Mathematics | **Mathematics** | Mathematics | Mathematics |
| Science | Science | **Science** | Science | Science | **Science** | Science |
| | ***Problem-solving*** | | | ***Problem-solving*** | ***Collaborative problem-solving*** | ***Global competence*** |

**Notes:** Major domain shown in bold; innovative domain in bold italics.

and sustainability education and all studies make increasing use of computer-based assessments. Applying the findings to inform changes in education systems is the key to improving learning, the ultimate goal for all of IEA's assessments.

### 3.2.2 PISA: Tracking learning outcomes and helping countries collect data on education[24]

Since 2000, the OECD's PISA has been providing internationally-comparable evidence on learning outcomes in reading, mathematics and science among 15-year-old students near the end of their compulsory education.

PISA assesses both student knowledge of subject content and students' capacity to apply that knowledge creatively, including in unfamiliar contexts. In each round of PISA, students are assessed in three core domains – reading, mathematics and science *(see **Figure 3.13**),* with one of these as the major domain in each cycle. In addition, one innovative domain – problem-solving, collaborative problem-solving or global competence, for example – is included in each cycle.

Confidence in the robustness of PISA is based on the rigour which is applied to all technical aspects of the survey design, implementation and analysis, not just on the nature of the statistical model, which continues to develop over time. Specifically on test development, the robustness of the assessment lies in the rigour of the procedures used in item development, conducting trials, analysis, review and selection. The task for the experts developing the assessment is to ensure that all these aspects are taken into account and to use their expert judgement to select a set of test items so that there is a sufficient balance across all these aspects. In PISA, this is done by assessment specialists who work with advisory groups made up of international experts. Participating countries and economies also play a key role in this item selection process.

PISA is conducted triennially to enable more than 80 countries *(see **Figure 3.14**)* to monitor their progress over time in meeting key learning objectives. The basic survey design has remained the same over the years to allow for comparability from one PISA assessment to the next and thus to allow countries to relate policy changes to improvements in education outcomes. By linking data on students' learning outcomes with data on key factors that shape learning in and outside of school, PISA highlights differences in performance patterns and identifies

24 Written by Michael Ward, Senior Analyst, Development Co-operation Directorate, OECD.

**Figure 3.14 Countries participating in PISA, 2015**


*Source: OECD.*

features common to high-performing students, schools and education systems.

Governments oversee decisions about PISA based on shared, policy-driven interests. New initiatives in the programme are considered in terms of their consistency with the programme's long-term strategy.

PISA is *a collaborative effort.* Decisions about the scope and nature of PISA assessments and the background information collected are undertaken by leading experts in participating countries.

### What does the evidence from PISA tell us?

The first results from PISA were published in December 2001, and they immediately sparked heated debate. The education landscape revealed by the assessment results was very different from what many had thought they knew. With each successive round of PISA, the results attracted more attention and triggered more discussion.

One of the most important insights from PISA is that education systems can change and improve. PISA shows that there is nothing inevitable or fixed about how schools perform. The results also show that there is no automatic link between social disadvantage and poor performance in school. These results have challenged everyone who thought education reform was impossible. If some countries can implement policies to raise achievement and narrow the social divide in school results, then why couldn't other countries be able to do the same?

In addition, some countries have shown that success can become a consistent and predictable education outcome. These are education systems where schools are reliably good. In Finland, for example, the country with the strongest overall results in the first PISA assessment, parents can rely on consistently high performance standards in whatever school they choose to enrol their child.

The impact of PISA is arguably greatest when the results reveal that a country performs poorly,

whether in absolute terms or in relation to a country's expectations. In these cases, PISA serves as a "wake-up call", raising public awareness and, in many cases, creating a momentum for reform. In an OECD 2012 survey of PISA-participating countries and economies, the large majority of respondents said that the policies of high-performing countries or improving systems had been influential in their own policymaking processes. The same number of countries and economies also indicated that PISA had influenced the development of new elements of a national or federal assessment strategy. In relation to curriculum setting and standards, many countries and economies cited the influence of the PISA frameworks on: comparisons of national curricula with PISA frameworks and assessments; formation of common standards nationally; impact on their reading frameworks; the incorporation of PISA-like competencies in their curricula; and for setting national proficiency standards.

The latest results from PISA 2015 tell us a great deal about quality and equity in education. Singapore outperformed all other participating countries and economies in science, the major domain in 2015. Japan, Estonia, Finland and Canada, in descending order of mean science scores, were the four highest-performing OECD countries. PISA describes six proficiency levels, with Level 6 the highest and Level 1 and below the lowest. Some 8% of students across OECD countries (and 24% of students in Singapore) were top performers in science, meaning that they were proficient at Level 5 or 6 on the PISA scale. Students at these levels are sufficiently skilled in and knowledgeable about science to creatively and autonomously apply their knowledge and skills in a wide variety of situations, including unfamiliar ones. About 20% of students across OECD countries scored below Level 2, considered the baseline level of proficiency in science. At Level 2, students can draw on their knowledge of basic science content and procedures to identify an appropriate explanation, interpret data and identify the question being addressed in a simple experiment. All students should be expected to attain at least Level 2 by the time they leave compulsory education.

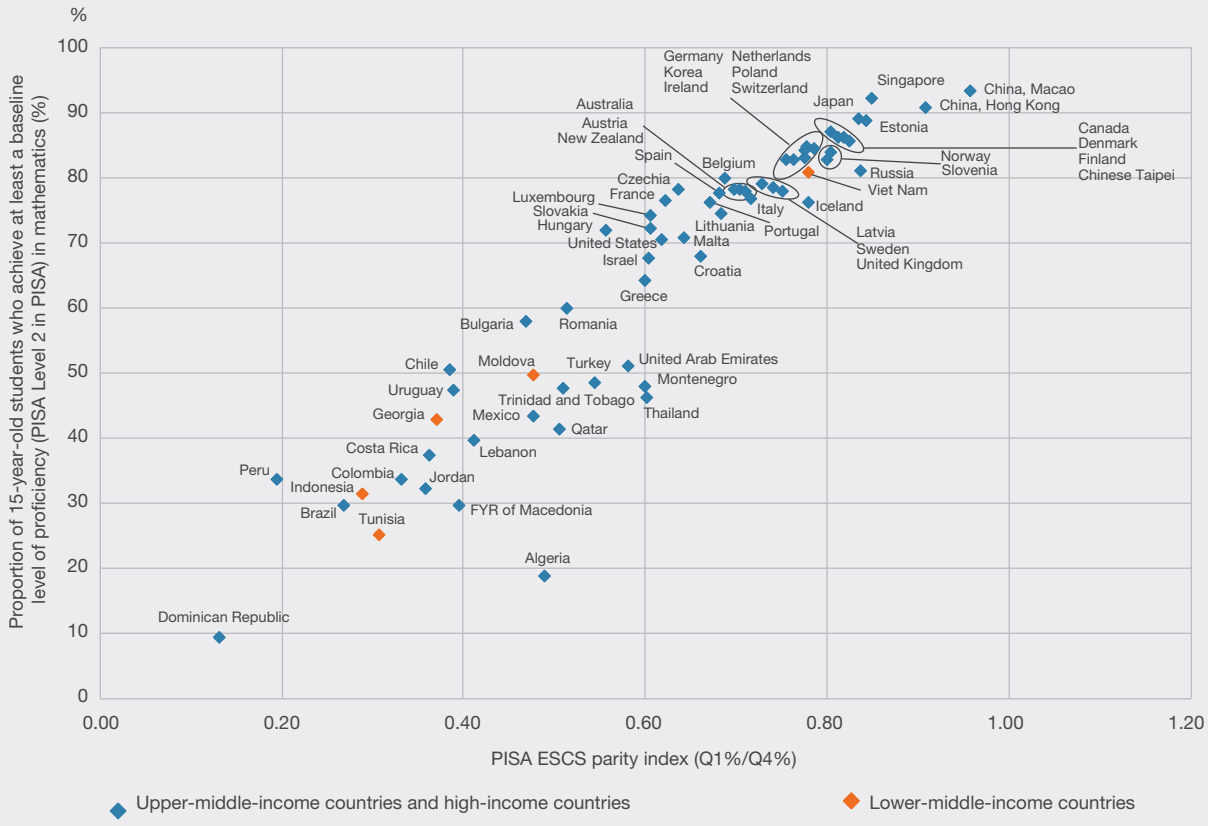
PISA 2015 also showed that Canada, Denmark, Estonia, Hong Kong Special Administrative Region of China and Macao Special Administrative Region of China achieve high levels of performance and equity in education outcomes. Socioeconomically-disadvantaged students across OECD countries were almost three times less likely than advantaged students to attain the baseline level of proficiency in science. However, about 29% of disadvantaged students were considered to be resilient – meaning that they beat the odds against them and performed at high levels. Additionally, in Macao Special Administrative Region of China and Viet Nam, students facing the greatest disadvantage on an international scale outperformed the most advantaged students in about 20 other PISA-participating countries and economies.

While between 2006 and 2015 no country or economy improved its performance in science and equity in education simultaneously, the relationship between socioeconomic status and student performance weakened in nine countries where mean science scores remained stable. The United States showed the greatest improvements in equity during this period. With regard to general improvement among OECD countries, average improvements (i.e. positive three-year trends) in reading performance between 2009 and 2015 were observed in Estonia, Germany, Ireland, Luxembourg, Norway, Slovenia and Spain. In mathematics, Albania, Colombia, Montenegro, Peru, Qatar and Russia improved their students' mean performance between 2012 and 2015, contributing to an overall positive trend since these countries began participating in PISA.

### PISA and SDG 4

SDG 4 has rightly shifted the focus from quantity (e.g. the number of children in school), which was a feature of the MDGs, to quality and equity. Quality (i.e. achievement) and equity (i.e. fairness and

**Figure 3.15 Proportion of 15-year-old students at the end of lower secondary education who achieve at least minimum proficiency in mathematics (PISA Level 2 or above)**




**Notes:** ESCS refers to the PISA index of economic, social and cultural status (see OECD, 2016a, for more information).
Parity is calculated as Q1%/Q4% where Q=Quartile of ESCS.
*Sources: Based on OECD, 2016b, PISA 2015 database, OECD, 2018 and World Bank, 2017b.*

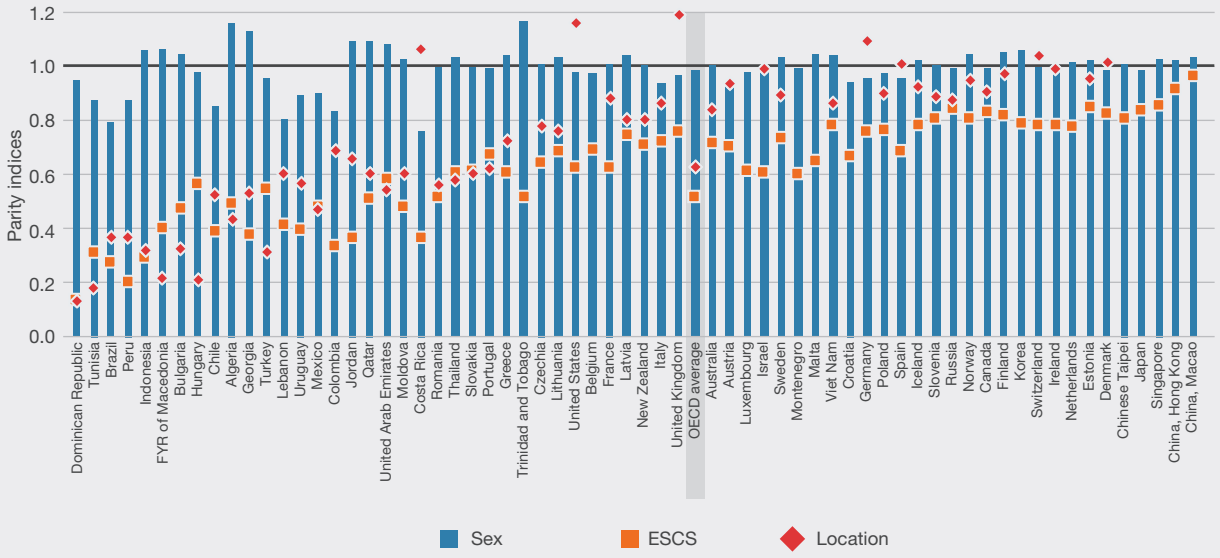
inclusiveness) are harder to measure than quantity; they require reliable, relevant and useful data on academic outcomes and participation. In order to serve the purpose of monitoring progress towards SDG 4, these data also need to be internationally comparable.

One of the global indicators selected to measure progress towards the first of the targets of SDG 4 (Indicator 4.1.1.c) is central to achieving quality education for all: the proportion of children and young people at the end of lower secondary education achieving at least minimum proficiency in reading and mathematics.

PISA has been identified by the UIS and the UN Statistical Commission (the two bodies responsible for monitoring progress towards SDG 4) as an internationally-comparable measure of this indicator. Since 2015, the UIS and the UN Statistical Commission have been using PISA data to report against global Indicator 4.1.1.c. PISA's proficiency Level 2 in reading and mathematics is considered to be the minimum level to be attained by students at the end of lower secondary education. Level 2 marks the baseline level of proficiency at which students begin to demonstrate the competencies that will enable them to participate effectively and productively in life as continuing students, workers and citizens.

**Figure 3.16 Sex, wealth and location parity index, 2015**

Proportion of 15-year-olds who achieve at least a PISA proficiency Level 2 in mathematics




**Notes:** Countries and economies are ranked in descending order of the sum of distance of each index to 1.
ESCS refers to the PISA index of economic, social and cultural status (see OECD, 2016a for more information).
Parity is calculated as Q1%/Q4% where Q=Quartile of ESCS.
*Sources: Based on OECD, 2016b, PISA 2015 database and OECD, 2018.*

As **Figure 3.15** shows, in Hong Kong Special Administrative Region of China, Macao Special Administrative Region of China and Singapore, at least 90% of students attain Level 2 or above in mathematics, while in Algeria, Brazil, the Dominican Republic and Tunisia, less than 30% of students attain this level of proficiency.

About 20% of students in OECD countries, on average, do not attain the baseline level of proficiency in reading. This proportion has remained stable since 2009. On average across OECD countries, the gender gap in reading in favour of girls narrowed by 12 points between 2009 and 2015; boys' performance improved, particularly among the highest-achieving boys, while girls' performance deteriorated, particularly among the lowest-achieving girls.

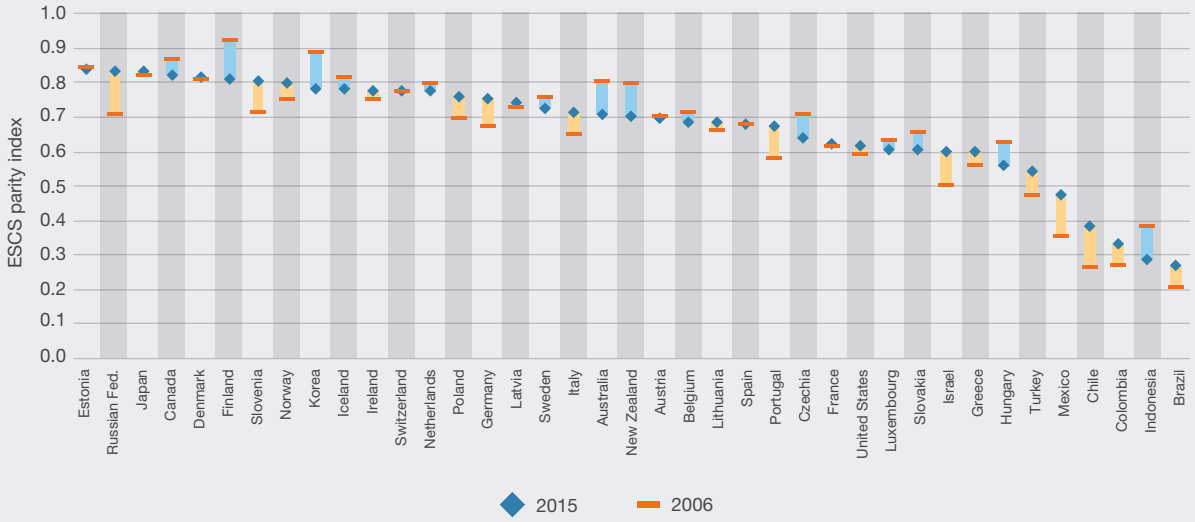
**Figure 3.16** shows parity indices[25] for Indicator 4.1.1.c by gender, location (urban or rural) and socioeconomic status (based on the PISA index of economic, social and cultural status [ESCS]). Among 15-year-old students, there are usually as many boys as girls who achieve at least proficiency Level 2[26] in mathematics, and more girls than boys who achieve Level 2 in reading. However, in the majority of OECD and partner countries, their performance remains strongly determined by their school's location. Students who attend urban schools (located in communities with over 100,000 inhabitants) are more likely to outperform those who attend rural schools (located

25 The parity index is defined as the ratio between the values of a given indicator for two different groups, with the value of the likely most disadvantaged group in the numerator. A parity index equal to 1 indicates parity between the two groups considered. A value less than 1 indicates a disparity in favour of the likely most advantaged group and a value greater than 1 a disparity in favour of the most disadvantaged group.

26 Although boys and girls are equally likely to perform at PISA Level 2 in mathematics, the gender gap in favour of boys widens at higher levels of performance.

**Figure 3.17 Trends in socioeconomic parity, 2006 and 2015**

Proportion of 15-year-olds who achieve at least PISA Level 2 in mathematics




**Note:** Countries are ranked in descending order of the ESCS parity index value in 2015.
*Source: OECD, 2018.*

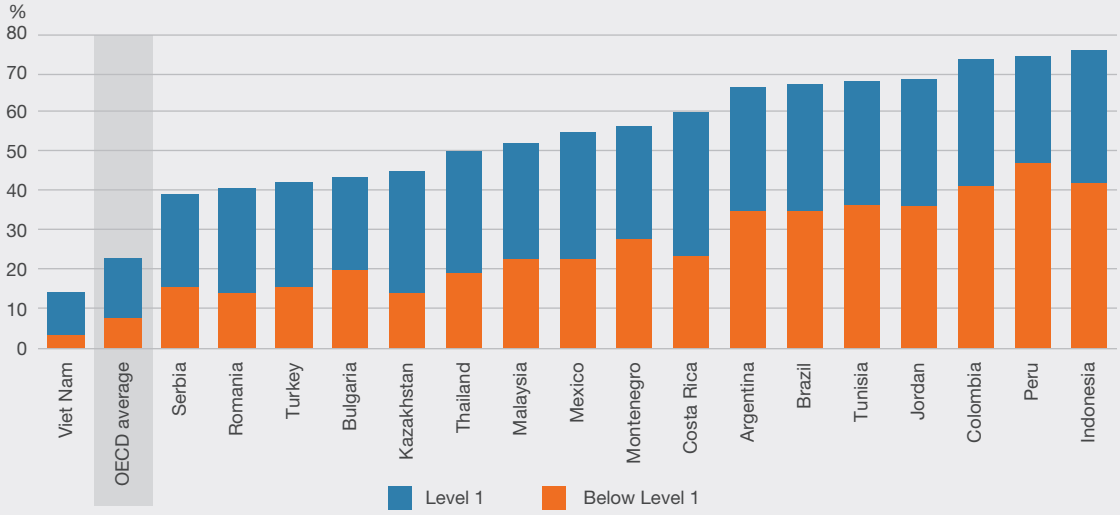
in communities with fewer than 3,000 inhabitants). Urban students tend to perform better because they go to schools that are usually larger and more likely to attract a larger proportion of qualified teachers. They are also more likely to come from a socioeconomically-advantaged background, which is directly linked to their performance in PISA (OECD, 2013a).

The performance gap between students from different socioeconomic backgrounds remains a reality in all countries, both in reading and mathematics. Even in those countries where parity is (almost) met along the three dimensions, such as Denmark, Slovenia and Estonia, the proportion of young people achieving PISA Level 2 in mathematics is 20% smaller among the most disadvantaged students. Even more worrying, levels of socioeconomic inequality have not changed since 2006 in the majority of countries.

**Figure 3.17** shows that in a few countries, including Australia, Finland and the Republic of Korea, the disparity between students in the top and bottom quartiles of the PISA index of socioeconomic status grew even larger between 2006 and 2015. However, PISA results show that inequity of opportunity is not set in stone and that selected school systems succeeded in becoming more equitable over a relatively short period (OECD, 2017). This is the case in Argentina, Mexico and the Russian Federation, where the performance gap between the quartiles of socioeconomic status narrowed significantly between 2006 and 2015. However, large differences in performance between disadvantaged and advantaged students remain in these countries.

PISA shows that in many countries, no matter how well the education system performs as a whole, socioeconomic status continues to predict student performance. However, PISA also consistently shows that high performance and greater equity are not mutually exclusive. Being able to improve the performance of all students, regardless of their background, is necessary for countries to become high performers and attain the SDG 4 targets.

**Figure 3.18 Percentage of students scoring at Level 1 or below in mathematics in 18 low- and middle-income countries, PISA 2012**




**Note:** Countries are ranked in ascending order of the percentage of students scoring below PISA Level 2 in mathematics.
*Source: OECD, 2014.*

## PISA for development

PISA results highlight differences in educational quality between high-income and middle-income countries: students in middle-income countries perform well below the OECD average *(see **Figure 3.18**)* and their performance is concentrated at the lower levels of the PISA proficiency scales. The limited differentiation of performance at these lower levels constrains the knowledge and understanding of what these students can and cannot do. It also limits the analyses that can be done, linking lower levels of learning with education policies and student characteristics.

Some of the contextual factors measured by PISA are unrelated to differences in performance in middle-income countries and PISA does not adequately reflect some of the contexts unique to these countries (Lockheed, Prokic-Bruer and Shadrova, 2015). Because many 15-year-olds in middle-income countries do not attend school, coverage can be as low as 50% (Spaull, 2017). In addition, some middle-income countries have encountered financial, technical and institutional difficulties in implementing the assessment and using PISA data.

PISA for Development (PISA-D) is making PISA more accessible and relevant to a wider range of countries. It is extending the PISA test instruments to measure a broader spectrum of performance, particularly at Level 2 and below. This is facilitating greater knowledge and understanding of what lower-performing students can do. It is developing contextual questionnaires and data collection instruments to capture the diverse situations in low- and middle-income countries. In addition, PISA-D is establishing methods and approaches to include out-of-school youth in the assessments – thus potentially offering a continuum between PISA and the OECD's Survey of Adult Skills (PIAAC) in terms of target populations and contributions to global SDG indicators, and it is building capacity in the participating countries to manage and use the results of large-scale student assessments.

While the PISA-D test design and items target the lower levels of performance, the assessment is linked

to the whole PISA framework for comparability. This link is established using data from some of the PISA 2015 trend questions. PISA-D provides a way of measuring differences in performance at the low end of the proficiency scale for each subject tested (reading, mathematics and science) even as it measures performance at the higher levels. The PISA-D cognitive test lasts two hours, as does the main PISA test, and the assessment is conducted in accordance with PISA's technical standards.

PISA-D participating countries (Bhutan, Cambodia, Ecuador, Guatemala, Honduras, Panama, Paraguay, Senegal and Zambia) were invited to join the project based on their experience and interest in large-scale assessments. The project is implemented with the support of a wide range of development and technical partners. The first results of the project will be released in December 2018 and the final results will be available in December 2019.

There is already evidence showing that the PISA-D project has helped build the capacity of the participating countries to manage and make good use of large-scale assessments. With the enhanced instruments and approaches from PISA-D made available in the main PISA test, it is possible that as many as 100 countries and economies may participate in the 2021 cycle of PISA. The project is on track to providing important insights into quality and equity in education in the participating countries, and to allow more countries to participate in PISA – all of which will help measure global progress towards achieving SDG 4 without excluding out-of-school youth.

## 3.3 MONITORING LEARNING OUTCOMES IN GPE DEVELOPING COUNTRY PARTNERS[27]

The GPE brings together Developing Country Partners (DCPs) (henceforth referred to simply as "partner countries"), donor nations, multilateral development organizations, civil society, teacher organizations, foundations and the private sector around a single shared vision: to ensure inclusive and equitable quality education and promote lifelong learning for all. The partnership aims to address educational challenges in some of the world's most demanding contexts. In 2002, only seven developing countries were GPE members. The number of partner countries has increased to 67 in 2018, with 32 of them being fragile or conflict-affected (FCAC) states. The number of partner countries is expected to continue growing as eligibility to join the partnership was extended to a total of 89 countries in 2017. Between 2002 and 2017, GPE cumulative grant allocation to partner countries amounted to US$ 4.8 billion (GPE, 2018).

In 2015, GPE adopted a strategic plan (GPE 2020) aiming to ensure inclusive and equitable quality education for all partner countries. This strategic plan covers the period 2016 to 2020. The GPE theory of change (GPE, 2017a) provides a framework of actions that can be undertaken at different levels (donors, developing country partners, local education groups, etc.) to strengthen education systems and promote inclusive and equitable quality education. A results framework that includes 37 indicators was introduced to monitor the Partnership's progress towards the GPE 2020 goals. One of the most important indicators in the GPE Results Framework is designed to monitor improvement in learning outcomes (Indicator 1).[28] The GPE Results Framework also monitors the status of learning assessment systems (Indicator 15) and the support that GPE provides to strengthening learning assessment systems in partner countries (Indicator 20).

Overall, learning outcomes are improving in GPE partner countries but the quality of the systems in place to monitor learning remains a significant challenge. While only two-thirds of GPE partner countries are expected to have conducted at least one learning assessment in the period 2016 to 2019,

27 Written by Élisé Wendlassida Miningou, Education Economist, and Ramya Vivekanandan, Senior Education Specialist, Global Partnership for Education.

28 Indicator 1 captures the percentage of partner countries showing improvement on learning outcomes in basic education (GPE, 2017c).

this is significant progress compared to a decade ago. Looking ahead, the challenges are to ensure that all partner countries have stable and sustainable learning assessments which are actually used to monitor and improve learning. To ensure that this happens and to strengthen the capacity of the learning assessment systems, GPE provides substantial support (financial and non-financial) to partner countries.

### 3.3.1 Learning trends in GPE partner countries

Through the learning outcomes improvement indicator (Indicator 1), GPE monitors learning trends over time using national, regional and international large-scale learning assessments. This indicator captures the proportion of partner countries showing improvement in learning outcomes in basic education from 2016 to 2020, the baseline period being 2000-2015. To inform the learning indicator, data must meet three key criteria: i) the data must be representative of the student population (including boys and girls) at either the national or sub-national level; ii) the learning assessment must measure achievements in language, mathematics and/or other key subject areas in basic education; and iii) the data must include learning level scores that are comparable across years (same subjects, same scale and drawn from equivalent samples of students) when more than one data point is available. When comparable data from two or more assessments are available, data are aggregated at the country level to compute the indicator value for the country.

A look at the baseline data from 20 partner countries with at least two data points available for the period 2000 to 2015 suggests reasonable progress.[29] In total, 13 countries – 65% of those with data – showed improvements in comparable learning assessments during the given timeframe.[30] Fewer countries affected by fragility and conflict (two of four in total) showed improvements.

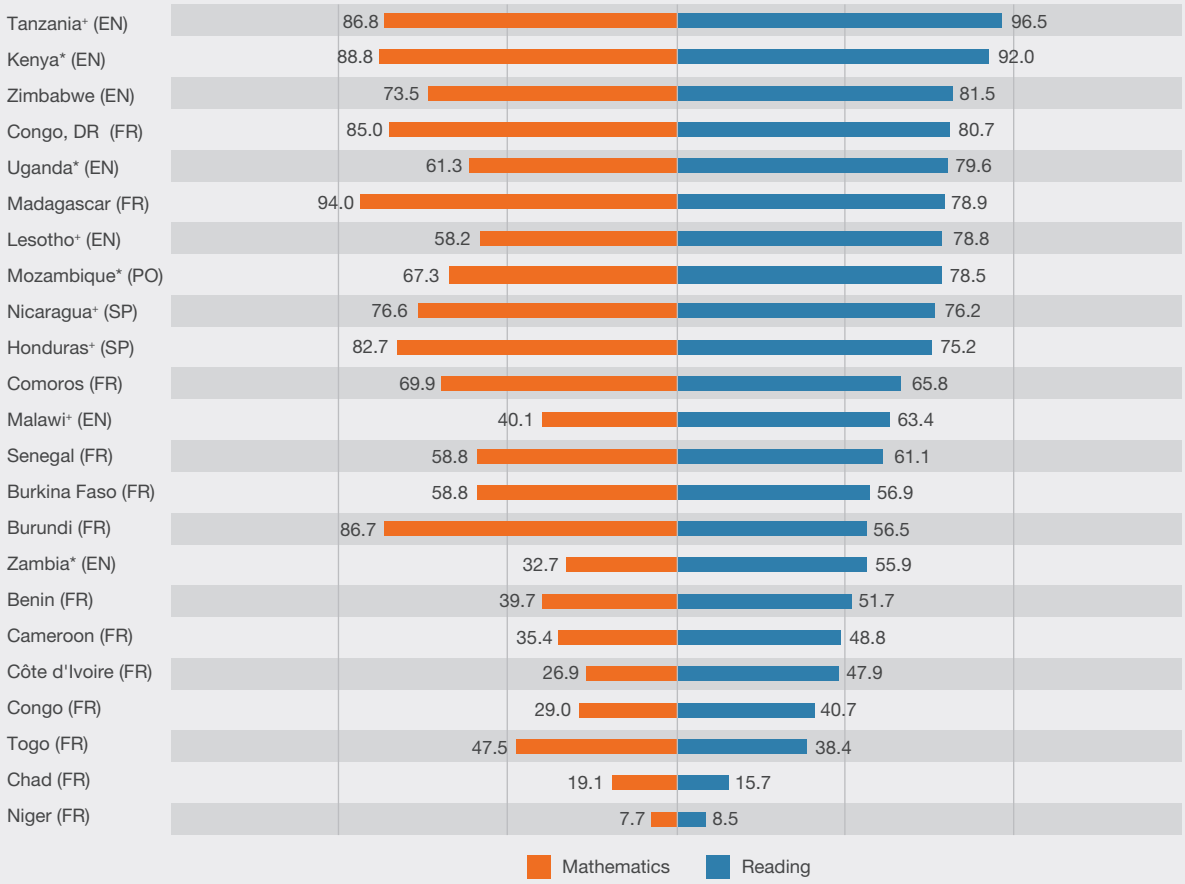
Despite the overall progress in learning, **Figure 3.19** shows results regarding the "absolute" level achievements in mathematics and reading. While in the United Republic of Tanzania 96% of students who completed primary education achieved the minimum proficiency level in reading, only 56% did so in Zambia. In mathematics, Madagascar is among the best-performing countries, while Niger is one of the countries with the lowest performance. Countries such as Niger and Chad performed relatively poorly in mathematics and reading, with less than 20% of students who completed primary education meeting the minimum proficiency level. While inter-subject comparison is not possible given the way that the assessments are calibrated, one can observe that overall achievement in mathematics is lower than achievement in reading in most countries.

According to the GPE Results Framework, some partner countries with relatively low achievement did not improve learning outcomes over the period 2000 to 2015 *(see Figure 3.19).* This shows that meeting the SDG Target 4.1 on minimum proficiency levels in reading and mathematics would require urgent action in these countries. For instance, in Zambia only 56% and 33% of students who completed primary education achieved the minimum proficiency levels in reading and in mathematics, respectively; overall, learning outcomes did not improve over the period 2000 to 2015.

Yet, at the same time some countries with relatively high achievements experienced improvement in learning outcomes between 2000 and 2015. For instance, in the United Republic of Tanzania and Lesotho, 96% and 79% of students who completed primary education achieved a minimum proficiency level in reading, respectively; learning outcomes improved over the period 2000 to 2015 in the two countries.

29 In 2015, 61 countries were GPE partners, and data on learning trends were only available in 20 of these countries.
30 These 13 countries include: Albania, Cambodia, Ethiopia, Georgia, Ghana, Honduras, Kyrgyzstan, Lesotho, Malawi, Moldova, Nicaragua, United Republic of Tanzania and Yemen.

**Figure 3.19 Percentage of students achieving at least a minimum proficiency level in reading and mathematics at the end of primary education, most recent data available between 2007 and 2015**



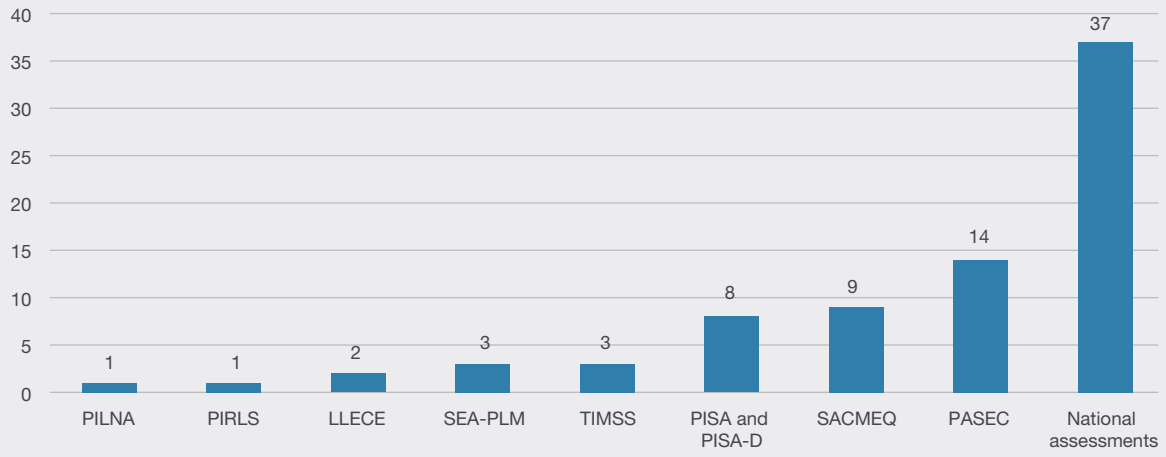

**Notes:** * Learning outcomes did not improve between 2000 and 2015.
+ Learning outcomes improved between 2000 and 2015.
Main language of instruction: EN: English; FR: French; SP: Spanish; PO: Portuguese.
DRC: Democratic Republic of the Congo.
Data are collected from regional and international learning assessments, such as PASEC, SACMEQ, LLECE and PISA. According to the UIS, the stardards for setting mininum proficiencies may vary by learning assesment and may not be comparable across learning assessments.
*Source: GPE compilations based on UIS and GPE Results Framework Indicator 1.*

## 3.3.2 Availability of learning assessment data in GPE partner countries

The availability of learning assessment data is a central challenge in reporting on learning outcomes. About 47 out of the 65 partner countries[31] have conducted a large-scale learning assessment over the period 2011 to 2015 that meets the criteria to inform the GPE learning trend indicator. However, based on information collected in February 2018, a total of 78 assessments are expected to take place in 48 GPE countries between 2016 and 2019 *(see **Figure 3.20**)*. While some countries already administer or will administer only one of these assessments, others will administer two or more (GPE, 2018). Nearly one-half of these will be national learning assessments (48% or 37 out of 78), followed by regional assessments (37% or 29 out of 78) and

31 Since the time that these data were collected, an additional two countries (Cabo Verde and Myanmar) have joined the Partnership.

**Figure 3.20 Number of learning assessments expected in partner countries between 2016 and 2019** [35]




*Source: GPE compilation based on publicly available information collected in February 2018.*

international assessments (15% or 12 out of 78). It is worth noting that nearly one-third (19 out of 67) of the partner countries are not expected to administer a large-scale learning assessment by 2019. This is particularly problematic in light of the SDG 4 targets, particularly Target 4.1 on minimum proficiency levels in reading and mathematics. If countries have no assessment in place to monitor learning levels, then they cannot know whether students are learning at minimum proficiency levels or learning at all.

Monitoring learning trends requires comparable data from frequently-conducted learning assessments that are representative at the national level (or at the regional or provincial level at least) and are specifically designed for monitoring learning outcomes over time. Data from examinations that are implemented for certification or selection purposes may not accurately capture learning trends for several reasons. In most cases, these examinations are not representative of the student population. In addition, given that examinations are sometimes used for filtering purposes (e.g. to determine progression to higher grades or levels of education), these are inherently not designed for inter-year comparability or, in other words, for monitoring learning trends over time.

Among the 48 partner countries that are expected to have carried out a large-scale learning assessment between 2011 and 2019, only a few are expected to have administered learning assessments that are comparable over time and therefore could be used to compute learning trends. It is expected that 43 partner countries will have the same learning assessments with at least two points in time (one assessment between 2011 and 2015 and another between 2016 and 2019). Of these 43 countries, learning assessments would be comparable in 31 countries. Reasons for non-comparability include changes in the assessment methodologies, variations in grades tested, variations in subject areas or changes in the assessment metrics. In other terms, monitoring the improvement of learning outcomes before and after the implementation of GPE 2020 would only be possible in about one-half of the partner countries.

There are several barriers to administering regular and comparable learning assessments in partner countries, and weak systems for learning assessments is one of these. To monitor the status

32 Some countries may administer more than one learning assessment.

of learning assessment systems, which refer to the overall ecosystem in which learning assessments are implemented, Indicator 15 from the GPE Results Framework considers the proportion of GPE partner countries which have a learning assessment system within the basic education cycle that meets quality standards (GPE, 2017d). The data around this indicator, as well as a summary of the challenges that partner countries face in building and maintaining such systems, is summarised in the section below.

### 3.3.3 Challenges with learning assessment systems in GPE partner countries

GPE's construct of a learning assessment system that meets quality standards, as defined for the purpose of the baseline data collection for Indicator 15 in 2016, looks specifically at large-scale assessments (including national, regional and international assessments as relevant) and public examinations. The construct is based on the World Bank's Systems Approach for Better Education Results (SABER) Student Assessment Framework but is further contextualised for the realities of GPE partner countries. The following aspects of these assessments are considered:

- Whether they have been carried out at regular frequency with all eligible students;
- Whether a permanent agency/institution/office is responsible for conducting the assessments;
- Whether the assessments are based on official learning standards or curriculum;
- Whether there are publicly available technical documents on the assessments;
- Whether the results are disseminated within a reasonable timeframe; and
- Whether assessment data are used to monitor learning outcomes (GPE, 2017b).

Using these criteria to analyse the different assessments in use in GPE partner countries, the indicator assigns a composite index to each country, which allows for classification of the overall system as "Established", "Under Development" or "Nascent". An assessment system meets the quality standards when it is classified as "Established".

The baseline data collected for this indicator in 2016 reveal that only 32% of all partner countries (19 out of 60) and only 21% of fragile and conflict-affected countries in the partnership were classified as "Established". Milestone data for Indicator 15 will be conducted this year (2018), while aspirational targets in this area are set for 2020.

The reasons that so many partner countries are struggling in this area are complex and manifold, as they face a host of technical, financial and political barriers in establishing and maintaining strong learning assessment systems. The technical challenges can include a lack of trained technical experts who can design, administer and analyse data from assessments. Teachers and school leaders, who are often responsible for administering assessments and then also on the receiving end as potential users of data from assessments, often lack assessment literacy. The quality of the assessment itself can also be a challenge; in many partner countries there are concerns about the validity and reliability of assessment items and instruments, as well as issues with sampling, test administration, data collection, cleaning and analysis.

Financial constraints can also be significant given that learning assessment can be expensive, with an average of US$500,000 needed per large-scale assessment for data collection and technical assistance, though there are cost differences between regions and depending on the size and complexity of the assessment (UIS, 2017e). In terms of the breakdown of these costs, the UIS reports that on average 20% of the budget of a national assessment is devoted to the assessment design and piloting, while 50% goes to the main test administration, 15% to data processing and analysis, and 15% to communication and dissemination (UIS, 2017d). In many developing countries, both government funding and support from development partners is insufficient to run these programmes at regular intervals.

It is nonetheless important to note that the costs of large-scale learning assessments as noted here represent a very small fraction of the average partner country's education spending. As noted in analysis from the UIS, the proportion of funds that countries expend on assessments, expressed in total costs per year per student, are minimal in comparison to overall government expenditure per student (UIS, 2016). Therefore financial constraints, while important, should not be considered a prohibitive barrier.

Even when adequate data on learning outcomes exist, they are often not used in a way that supports improvements in learning. The reasons for this vary. Policymakers are often not involved in the assessment cycle, whereas their engagement and related identification of key policy challenges and questions, and integration of these in the assessment design, are crucial to ensuring that assessment results are actually used. The results of learning assessments are often presented in inaccessible language and formats that may not be tailored to the needs of different stakeholders, including policymakers, teachers, parents and students. In other cases, the results are simply not disseminated at all, particularly to school-level stakeholders. This can be observed particularly in cases when the assessment in question shows poor results. Sharing such results can be politically risky, calling into question the overall performance of the public education system and those who are responsible for it. Even when assessment data are disseminated and provide a basis on which policies can be designed and adjusted, the resources to make necessary changes can be inadequate or the data may not reach those who can actually make the changes. In addition, in order to be useful, data on learning outcomes should be analysed in concert with contextual factors to determine how different groups of students are learning. Often, all kinds of data are collected, but they are either not sufficiently analysed or not specifically linked to contextual factors which can inform how to intervene.

Also, very crucially, there are often different expectations or a lack of alignment and coordination between assessment, curriculum and pedagogy. The units which are responsible for these three domains may not have a sense that they are responsible to each other or that they should be supporting each other. In some cases, this situation is compounded by the very organizational design of ministries of education, which may keep these units apart and not accountable to one another. Yet if the articulation between assessment, curriculum and pedagogy is not present, the ability for countries to use data to drive improvements in education can be vastly compromised.

Addressing these challenges to strengthen learning assessment systems is a key priority for the GPE. It is hoped that the partnership's efforts in this area, as summarised in the next section, will contribute to this.

### 3.3.4 GPE support to monitoring and improving learning for all

Given that the GPE is committed to the strengthening of learning assessment systems in its partner countries, it is useful to consider the ways in which the partnership provides this support through the GPE fund. From a macro perspective, GPE's Results Framework includes an indicator (Indicator 20) that examines the proportion of its country-level grants that support EMIS and/or learning assessment systems. Using data collected for Indicator 20, it is possible to determine the number of DCPs receiving GPE support towards this goal. As of 30 June 2017, 29 of the 41 active Education Sector Program Implementation Grants (ESPIGs) were investing in learning assessment systems.[33]These grants support activities such as the development and implementation of classroom-based and national assessments in Bangladesh, the establishment of an independent agency in charge of national assessments in the Democratic Republic of Congo, and the support given to Sudan described in **Box 3.1**.

33 This number does not include sector-pooled grants.

## Box 3.1 Building systems for teaching and learning data in Sudan

Sudan joined the GPE in 2012, following a political crisis that left over 2 million people internally displaced. With no system to collect basic education data on service delivery and learning outcomes, the government committed to building capacity to collect, analyse and utilise data for educational planning and system-wide improvements.

Sudan received a GPE grant of US$76.5 million to assist in the implementation of the Basic Education Recovery Project, which focuses on improving the learning environment for basic education and strengthening education management and planning. The GPE project supports the establishment of a National Learning Assessment (NLA), which in 2015 rolled out a nationwide learning assessment across 18 states, involving approximately 10,000 students in over 450 schools. The assessment was aimed at gaining an understanding of literacy and numeracy at the end of Grade 3, which corresponds to the end of the first cycle of basic education, using a modified Early Grade Reading Assessment (EGRA).

*Source: GPE, 2016.*

In addition, the GPE funding model (adopted in 2014 by the Board for the 2015-2018 replenishment period) requires countries applying for a Program Implementation Grant to either have a system in place to monitor learning outcomes or a plan to develop one. The GPE financing is expected to provide support in this regard.[34]The funding model also allows GPE to provide results-based financing, which gives countries incentives to set and achieve their own learning targets. To receive the first 70% of GPE funding, each DCP must meet several key requirements, including having a system or mechanisms in place to monitor learning outcomes, as explained above. Disbursement of the remaining 30% is linked to demonstrated progress toward sector results, including gains in learning. Governments, in consultation with their partners in the local education group, must identify a transformational strategy to improve learning outcomes that outlines clear actions to remedy issues driving low learning levels in their context. For example, the Democratic Republic of the Congo has linked funding to improved reading performance in the primary grades.

The GPE has also supported global and regional activities to strengthen learning assessment systems. The Global and Regional Activities (GRA) programme, which has formally concluded, contained two grants focused on learning assessment systems. The first grant supported a UIS project from 2013 to 2015 to develop methodologies to link reading assessments across regions, identify best practices for early reading assessment and initiate a global catalogue of learning assessments. The second grant supported the initial activities of the Network on Education Quality Monitoring in the Asia-Pacific (NEQMAP) to build regional evidence and capacity. These activities were implemented by UNESCO Bangkok from 2014 to 2016.

Following these activities, and as a pilot for the Knowledge and Good Practice Exchange (KGPE) approach, the Assessment for Learning (A4L) initiative was launched in July 2017 to build capacity for national learning assessment systems to measure and improve learning. A4L has three components: i) tools to support diagnostics of learning assessment systems, to be made publicly available after piloting in three DCPs in 2018 and 2019; ii) support to NEQMAP in the Asia-Pacific region (coordinated by UNESCO Bangkok) and the Teaching and Learning Educators' Network for Transformation (TALENT) in sub-Saharan Africa (coordinated by UNESCO Dakar) for capacity development, analytical work and knowledge sharing; and iii) a landscape review of measurement of 21st century skills and tools to support such measurement (with the Center for Universal Education at the Brookings Institute).

34 Public examinations and issuing diplomas do not count toward this requirement, unless specifically used to monitor learning outcomes.

Looking towards the future, strengthening learning assessment systems will be a priority thematic area for GPE's Knowledge and Innovation Exchange (KIX), a new initiative to support the development of global and regional goods and sharing of knowledge and expertise. As such, GPE will be able to support an expanding portfolio of work in this area. In concert, these various GPE efforts to strengthen learning assessment systems aim to ultimately contribute to the improvement of learning outcomes in partner countries. Continued progress is expected on this front over the period of GPE 2020 and beyond.

These improvements and the ability to measure and report on the learning levels of children are crucial to support the efforts to monitor SDG Target 4.1 and Indicator 4.1.1. The international community, through GAML coordinated by the UIS, is working hard to better define this indicator and support the efforts of countries to monitor and report on it. The work of GPE, as summarised in this section, is expected to make a significant contribution.

## 3.4 EGRA AND EGMA: UNDERSTANDING FOUNDATIONAL SKILLS[35]

The Early Grade Reading Assessment (EGRA) was commissioned in 2006 by the U.S. Agency for International Development (USAID) to assist low-income countries in rapidly diagnosing and improving early reading outcomes (RTI International, 2007). While one donor and partner took this initial concrete step, other actors, such as Pratham in India, were thinking along the same lines. Designed with similar intent, the Early Grade Mathematics Assessment (EGMA) was established shortly after EGRA (RTI International, 2009).

EGRA and EGMA are individually-administered oral assessments of foundational skills that are predictive of future performance. For EGRA, the skills assessed include letter and/or letter-sound identification, phonemic awareness, familiar word reading and oral reading fluency. EGMA includes basic counting, number identification and number patterns, magnitude (number discrimination) and simple operations (addition and subtraction). Some applications of EGRA or EGMA include other tasks, but those noted are included in most applications and are generally considered the most important. Assessors guide students through each of the sub-tasks, providing detailed instructions and practice items. Students read from a paper stimulus, while assessors score responses on a digital tablet or paper form. Test sections include timed and untimed portions, with rules for discontinuing the task if students fail to respond to the first few items correctly.

The purpose of the assessments is to inform stakeholders about the strengths of and gaps in teaching and learning in the early grades of primary school (Dubeck and Gove, 2015; Platas, Ketterlin-Geller and Sitabkhan, 2016). Open-source toolkits guide the development, administration and analysis of results for each language and country-specific assessment (Platas et al., 2014; RTI International, 2016). Donor support and an open-access approach to sharing instruments and datasets have accelerated the use of the tools, particularly in the Global South. We estimate that EGRA has been administered in more than 70 countries and in 120 languages (RTI International, 2016). Support for and expansion of EGMA has lagged behind EGRA, because USAID programmes have prioritised reading. As of the last formal assessment, EGMA has been applied in approximately 25 countries.

### 3.4.1 What have we learned so far from EGRA and EGMA results?

EGRA and EGMA results have been used to inform system-level policy and programme-level impact. Results from the reading assessment indicate that low levels of learning are widespread among children in the early grades of primary school. For several low-income countries (or the more vulnerable regions of the countries), more than 90% of students enrolled in Grade 2 are unable to read a single word of a grade-

35 Written by Luis Crouch, Senior Economist, and Amber Gove, Director of Research, RTI International.

level reading passage (Gove et al., 2015). **Figure 3.21** presents a summary of EGRA fluency distributions in national-level studies for Grades 2 and 3. The distributions skew to the left, with the largest share of students in the lowest-performing fluency segment, zero to ten words read per minute. For this group of countries, Mali recorded the highest proportion of Grade 2 students in the lowest fluency segment, while Rwanda has the smallest. Reviewing the distributions, as opposed to means, is important for understanding where to best channel resources to raise overall learning levels. The distributions suggest the need to invest in strategies to support the large groups of low performers.

These results, with evidence from similar approaches to understanding early skills development (UIS, 2016), informed the design of SDG 4. Though the availability of data on early learning was not the only factor in shifting the focus from access to learning, the recognition of the scale of the learning crisis contributed to an increased emphasis on learning relative to prior frameworks of global goals. In particular, Indicator 4.1.1a calls on countries to measure and report on the proportion of children in Grades 2 or 3 reaching at least a minimum proficiency level in reading and mathematics.

Countries have the option of using nationally-representative EGRA and EGMA outcomes to report on this indicator, although other measures also exist which meet international standards of quality. Results for this indicator, shown in **Table 3.4**, are estimated by using the percentage of students reaching country-specific reading and mathematics proficiency standards developed with ministries of education. These standards are typically country-specific because written languages differ in their complexity and, for the early grades, it is reasonable to set different expectations for different languages. The data reported here should be considered carefully, as countries are solely responsible for reporting on this indicator to the UIS.

In addition to identifying reading and mathematics proficiency levels, the EGRA and EGMA experience has also provided insight into how countries might monitor issues at foundational levels of education. Identifying shortfalls at a young age can inform policy and encourage corrective action. EGRA and EGMA methods are relatively easy to understand and apply, as the tasks used, while based on the science of reading and mathematics education, also reflect a layperson's understanding of what early literacy and numeracy mean.

In addition to reporting against the SDG indicator, another way to approach multi-national comparisons is to report on the percentage of students unable to correctly read *any* of the items presented, known as zero scores. Since languages differ in the opacity of the scripts they use, most users of EGRA stay away from fluency comparisons across countries and, especially, across languages. Comparisons of the percentage of children unable to read any words *at all* would be less affected by language opacity. **Table 3.5** presents the proportion of zero scores on an oral reading fluency assessment administered to students in Grades 2 and/or 3 for 20 locations with several results presented at the regional level. While most countries report one year of results on a national scale, several countries have collected data from multiple years.

### 3.4.2 How are EGRA and EGMA results being used to monitor and support learning?

EGRA and EGMA results are frequently used to develop new approaches or programmes in countries. In Kenya, USAID supported the use of EGRA and EGMA to inform the implementation of the Primary Math and Reading (PRIMR) programme. Since early 2010, PRIMR supported the partnership between the Kenyan Ministry of Education and RTI International in the development and testing of a package of innovations to improve early reading and mathematics

**Figure 3.21 Distribution of oral reading fluency scores by grade and language for national samples, 2009-2013**


*Source: Gove et al., 2015.*

**Table 3.4 Percentage of students reaching minimum proficiency in reading**

| Place/language | All | Boys | Girls | Year | Proficiency indicator definition [36] * |
|---|---|---|---|---|---|
| **Ethiopia** | | | | | |
| Afaan Oromo | 5% | 6% | 3% | 2014 | > 48 cwpm |
| Af Somali | 14% | 16% | 13% | 2014 | > 50 cwpm |
| Amharic | 6% | 6% | -- | 2014 | > 50 cwpm |
| Hadiyyisa | 4% | 6% | 3% | 2014 | > 40 cwpm |
| Sidamu Afoo | 1% | 2% | 0% | 2014 | > 45 cwpm |
| Tigrinya | 0.30% | 0.70% | 0% | 2014 | > 55 cwpm |
| Wolayttatto | 8% | 9% | 7% | 2014 | > 43 cwpm |
| **Ghana** | | | | | |
| Ghanaian languages avg. | 3% | 3% | 2% | 2013 | 40 cwpm |
| English | 7% | 6% | 7% | 2013 | 45 cwpm |
| Jordan | | | | | |
| Modern Standard Arabic (MSA) | 3%* | 1% | 4% | 2014 | 46 cwpm |
| **Liberia** | | | | | |
| English | 4% | 6% | 3% | 2013 | 35-40 cwpm |
| **Malawi** | | | | | |
| Chichewa | 0.24% | 0.22% | 0.26% | 2012 | 40 cwpm |
| **Pakistan** | | | | | |
| Urdu | 20% | 16% | 25% | 2013 | 60-90 cwpm |
| Sindhi | 24% | 26% | 22% | 2013 | 50-80 cwpm |
| **Philippines** | | | | | |
| Ilokano | 35% | 27% | 46% | 2014/2013 | 40 cwpm (2014 data) |
| Hiligaynon | 34% | 29% | 40% | 2014/2013 | 45 cwpm (2014 data) |
| Cebuano | 54% | 41% | 62% | 2014/2013 | 42 cwpm (2014 data) |
| Maguindinaoan | 22% | 15% | 30% | 2014/2013 | 40 cwpm (2014 data) |
| **Tonga** | | | | | |
| Tongan | 15% | -- | -- | 2009 | 50 cwpm |
| **UR Tanzania** | | | | | |
| Kiswahili | 5% | 4% | 6% | 2013 | 50 cwpm |
| Vanuatu | | | | | |
| Eng. and French | 6% | 3% | 7% | 2010 | 45 cwpm |
| **West Bank** | | | | | |
| Arabic | 18% | 15% | 22% | 2014 | 30 cwpm with diacritics |
| Arabic | 27% | 33% | 21% | 2014 | 35 cwpm without diacritics |
| **Zambia** | | | | | |
| Avg. seven languages | 1% | 1% | 1% | 2014 | 45 cwpm |

36 www.dec.usaid.gov

## Table 3.5 Grade 2 or 3 oral reading fluency zero scores, by location and sex, 2010-2015

| Place | Year | Grade | Boys (%) | Girls (%) | All (%) |
|---|---|---|---|---|---|
| DR Congo – Equateur | 2015 | 3 | 77 | 86 | 81 |
| DR Congo – Kasai Occidental | 2015 | 3 | 71 | 89 | 80 |
| DR Congo – Kasai Oriental | 2015 | 3 | 75 | 86 | 80 |
| DR Congo – Katanga | 2015 | 3 | 85 | 84 | 84 |
| Egypt | 2013 | 3 | 25 | 18 | 22 |
| Ghana | 2013 | 2 | 84 | 82 | 83 |
| Iraq | 2012 | 2 | 37 | 31 | 34 |
| Iraq | 2012 | 3 | 21 | 12 | 17 |
| Jordan | 2012 | 2 | 25 | 17 | 21 |
| Jordan | 2012 | 3 | 24 | 16 | 20 |
| Jordan | 2014 | 2 | 16 | 7 | 11 |
| Jordan | 2014 | 3 | 7 | 2 | 4 |
| Liberia | 2013 | 2 | 26 | 36 | 31 |
| Liberia | 2013 | 3 | 9 | 13 | 11 |
| Malawi | 2010 | 2 | 94 | 95 | 94 |
| Malawi | 2012 | 2 | 89 | 91 | 90 |
| Mali – Classique | 2015 | 2 | 65 | 69 | 67 |
| Mali – Curriculum | 2015 | 2 | 59 | 57 | 58 |
| Morocco | 2011 | 2 | 36 | 29 | 33 |
| Morocco | 2011 | 3 | 22 | 13 | 18 |
| Nepal | 2014 | 2 | 35 | 39 | 37 |
| Nepal | 2014 | 3 | 18 | 20 | 19 |
| Nigeria – Bauchi | 2013 | 2 | 94 | 97 | 96 |
| Nigeria – Bauchi | 2013 | 3 | 83 | 86 | 84 |
| Nigeria – Jigawa | 2014 | 2 | 82 | 88 | 84 |
| Nigeria – Jigawa | 2014 | 3 | 80 | 87 | 83 |
| Nigeria – Kaduna | 2014 | 2 | 97 | 96 | 97 |
| Nigeria – Kaduna | 2014 | 3 | 85 | 90 | 88 |
| Nigeria – Kano | 2014 | 2 | 86 | 91 | 88 |
| Nigeria – Kano | 2014 | 3 | 70 | 78 | 74 |
| Nigeria – Katsina | 2014 | 2 | 88 | 89 | 88 |
| Nigeria – Katsina | 2014 | 3 | 74 | 83 | 78 |
| Nigeria – Sokoto | 2013 | 2 | 91 | 98 | 94 |
| Nigeria – Sokoto | 2013 | 3 | 84 | 95 | 88 |
| Philippines – ARMM – Maguindanaoan | 2014 | 2 | 45 | 30 | 38 |
| Philippines – ARMM – Maguindanaoan | 2015 | 2 | 32 | 25 | 29 |
| Philippines – Region I – Ilokano | 2014 | 2 | 16 | 9 | 13 |
| Philippines – Region I – Ilokano | 2015 | 2 | 15 | 4 | 10 |
| Philippines – Region VI – Hiligaynon | 2014 | 2 | 23 | 20 | 22 |
| Philippines – Region VI – Hiligaynon | 2015 | 2 | 29 | 19 | 25 |
| Philippines – Region VII – Cebuano | 2014 | 2 | 11 | 4 | 8 |
| Philippines – Region VII – Cebuano | 2015 | 2 | 7 | 1 | 5 |
| Philippines – English | 2013 | 3 | 2 | 0 | 1 |
| Papua New Guinea – East New Britain | 2012 | 2 | 25 | 19 | 22 |
| Papua New Guinea – East New Britain | 2012 | 3 | 13 | 11 | 12 |
| Papua New Guinea – Madang | 2011 | 2 | 12 | 18 | 15 |
| Papua New Guinea – Madang | 2011 | 3 | 7 | 5 | 6 |
| Papua New Guinea – National Capital District | 2012 | 2 | 38 | 38 | 38 |
| Papua New Guinea – National Capital District | 2012 | 3 | 13 | 16 | 14 |
| Papua New Guinea – Western Highlands | 2013 | 2 | 32 | 34 | 33 |
| Papua New Guinea – Western Highlands | 2013 | 3 | 13 | 8 | 11 |
| Tonga | 2009 | 2 | 13 | 4 | 9 |
| Tonga | 2009 | 3 | 3 | 3 | 3 |
| Uganda | 2015 | 2 | 66 | 62 | 64 |
| Uganda | 2015 | 3 | 40 | 30 | 35 |
| UR Tanzania | 2013 | 2 | 31 | 25 | 28 |
| Vanuatu | 2010 | 2 | 30 | 20 | 25 |
| Vanuatu | 2010 | 3 | 8 | 6 | 7 |
| West Bank | 2014 | 2 | 27 | 17 | 22 |
| Yemen | 2011 | 2 | 45 | 38 | 42 |
| Yemen | 2011 | 3 | 36 | 15 | 27 |
| Zambia | 2015 | 2 | 56 | 57 | 56 |

*Source: Early Grade Reading Barometer Comparisons Report, USAID, 2018.*

instruction. Components of this package included scripted teachers' guides, student textbooks and workbooks, teacher training and continuous coaching support.

As described in Piper and Mugenda (2014), EGRA and EGMA were used to track student performance at multiple levels. At the classroom level, teachers were trained on how to assess and track individual student progress throughout the school year and provide additional practice to those students who did not meet performance expectations. Instructional coaches also used EGRA to spot-check student reading performance and discuss results with teachers. Finally, the programme designers used EGRA and EGMA to test the relative effectiveness of multiple variations of the intervention, including the ideal ratio of coaches to teachers, whether teachers' guides should include lesson plans, and the use of the mother tongue as a medium of instruction to support learning in a second or third language. In this way, the assessment helped to inform continuous adaptation and improvements to the programme design, resulting in a national approach, called Tusome, that was scaled to all Grade 1 to 3 classrooms throughout Kenya's 25,000 public, private and alternative schools.

With USAID support, EGRA and, to a lesser extent, EGMA have been used to inform programme design and continuous improvement in dozens of countries, including the Democratic Republic of the Congo, Egypt, Guatemala, Indonesia, Liberia, Malawi, Mali, Mozambique, Nepal, Jordan, the Philippines, Rwanda, Senegal, the United Republic of Tanzania, Uganda and Zambia. Detailed reports from each of the USAID-funded data collections and programme interventions can be found on the Development Experience Clearinghouse website.[37] Summary results for multiple countries are also available via USAID's Early Grade Reading Barometer.[38]

### 3.4.3 What are the challenges to inform SDG 4?

**Comparability:** Since languages differ in their opacity, EGRA results, specifically in fluency, are not necessarily representative of the quality of an education system. However, if reports are based on countries' own expectations of child performance and take issues such as inherent script difficulty into account, cross-national comparisons may be possible.

**Quality assurance**: While applying EGRA is in many ways easier than applying traditional pencil-and-paper assessments closer to the end of the primary cycle, EGRA applications are still technically non-trivial. EGRA should never be translated; one generally wants to see a serious adaptation that takes into account the script and other opacity considerations of the language. Serious inter-rater reliability exercises have to be carried out. Validity and reliability need to be measured. Proper sampling is still necessary. Before accepting EGRA reports, the recipients of these reports would do well to assess the technical quality of the application, just as with any assessment.

**Collaboration and sharing of tools:** International organizations should encourage more assessment providers to openly share tool design, construction and capacity-building approaches. Comparisons of benchmarks would also help, so that one knows what is being compared to what. More could be done to popularise lessons learned from through the utilisation of oral assessment tools, such as UNESCO's Literacy Assessment and Monitoring Programme (LAMP), along with EGRA, ASER and similar assessments.

37 www.dec.usaid.gov
38 www.earlygradereadingbarometer.org

## 3.5 LEARNING EVIDENCE IN READING AND ARITHMETIC IN CHILDREN AGED 5 TO 16 YEARS IN INDIA[39]

### 3.5.1 Overview

Thanks to sustained policy focus for well over a decade, today almost all children of elementary school age in India are enrolled in school. In 2005, the first Annual Status of Education Report (ASER)[40] reported that 6.6% of all children in the official elementary school-age group of 6- to 14-year-olds in rural India were not enrolled in school, a figure that almost exactly matched the official estimate of 6.9% produced by a study commissioned by the Government of India (IMRB, 2014). ASER 2016 reported 11 years later that the proportion of unenrolled 6- to 14-year-olds had dropped by one-half, to 3.3%. With close to 97% children enrolled, the country currently has about 200 million children in elementary school (Classes 1 through 8), or about 25 million children per elementary grade, distributed over more than 1.4 million schools across the country (National University for Educational Planning and Administration, 2017).

Are children's learning outcomes also at satisfactory levels? The body of evidence on children's learning has grown in recent years. Today data are available from a range of sources, including large-scale learning assessments conducted by both government and non-government institutions, as well as research studies that have examined children's learning and its determinants.

However, the only current source of annual, comparable data available on scale in India is the annual ASER survey, first implemented in 2005. Over the years, ASER has provided annual estimates of basic reading and arithmetic for a sample of children aged 5 to 16 from an average of about 570 rural districts in India.[41] ASER is designed as a household-based, rather than a school-based, survey in order to ensure that all children are included rather than only those enrolled and present in school on the day of the survey.[42]

ASER employs a "floor" level test of basic reading and arithmetic: that is, the same test is administered to all children aged 5 to 16 regardless of age, grade or enrolment status. The assessment is administered orally, one-on-one (individual administration) with each sampled child.

The reading assessment tool consists of four simple reading tasks illustrated in **Figure 3.22**. The easiest task comprises reading letters of the alphabet, followed by simple commonly-used words. The third reading task comprises a paragraph with four short sentences, equivalent to text that children are expected to be able to transact in Class 1 of primary school. The most difficult task involves reading a slightly longer, more complex text equivalent to the contents of a Class 2 textbook. Tools are currently available in 20 Indian languages, including English, which covers the language of instruction in early grades of virtually all schools in the country. The arithmetic test has a similar design and contains four tasks: single-digit number recognition, double-digit number recognition, two-digit by two-digit subtraction with borrowing, and three-digit by one-digit division.

In both reading and arithmetic, younger children in Classes 1 and 2 are not expected to be able to go beyond the first couple of tasks. However, it is expected that from Class 3 onwards, children should

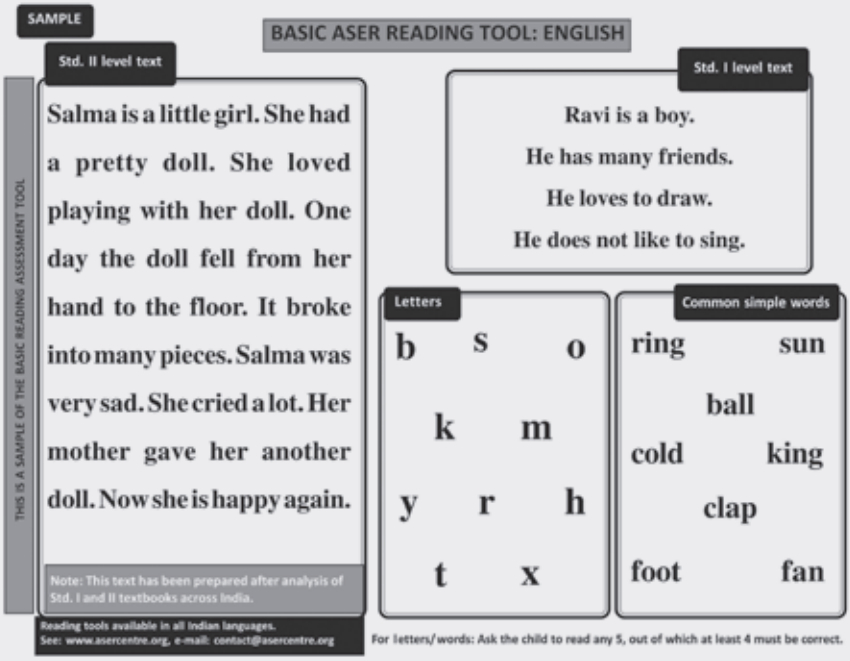
39 Written by Rukmini Banerji, Director, and Suman Bhattacharjea, Director of Research, ASER Centre.

40 ASER is an annual household-based assessment that generates estimates of schooling status for children age 3-16 and of foundational reading and arithmetic ability for children age 5-16. The learning assessment is administered one on one with each child. Estimates are representative at the district, state, and national level. Facilitated by the non-government organization Pratham and conducted by partner organizations in almost all of India's rural districts, the survey has reached more than half a million children each year since 2005.

41 ASER is designed to generate representative estimates at district, state and national levels. The survey employs a two-stage sample design, with villages being sampled in the first stage and households in the second stage. All children in the 3 to 16 age group in sampled households are surveyed, but only those aged 5 to 16 are tested.

42 Although only a small proportion of children in India is not enrolled in school, absenteeism is a major problem, with an average of about 30% students in Classes 1 to 5 being absent on a random day in the year. In some states this proportion is as high as 50%. Further, a growing proportion of children attend private schools which may be unrecognised and/or unaided, and may thus be missing from the official lists of schools. Generating a representative picture of all children therefore requires household-based sampling.

**Figure 3.22 The ASER reading assessment tool in English**




*Source: ASER India.*

be able to comfortably and confidently complete the simple tasks in the ASER assessment.[43]

### 3.5.2 Three broad trends

Broadly, three clear trends are illustrated in the ASER data from 2006 to 2016. First, children's foundational learning levels are low and remain low over time. This is the most frequently cited finding from ASER.

In 2006, ASER reported that 53% of all children enrolled in Class 5 across the country could read a simple text at a level of difficulty three grades below. In other words, even after four years of schooling, only slightly over one-half of all children were able to comfortably read a text at Class 2 level of difficulty, such as the text labelled "Std II level text" in Figure 3.22**.** This proportion did not increase over the following decade and in fact was observed to decline further after 2010. By 2016, just 48% of students in Class 5 were able to read a Class 2 level text *(see **Table 3.6**).*

**Table 3.6 Percentage of children from Classes 3, 5 and 8 who can read a Class 2 level text**

| Year | Class 3 (%) | Class 5 (%) | Class 8 (%) |
|---|---|---|---|
| 2006 | 20.0 | 53.1 | 83.8 |
| 2008 | 22.2 | 56.2 | 84.8 |
| 2010 | 19.5 | 53.7 | 83.5 |
| 2012 | 21.4 | 46.8 | 76.4 |
| 2014 | 23.6 | 48.0 | 74.6 |
| 2016 | 25.1 | 47.8 | 73.0 |

*Source: ASER India.*

43 The basic reading and arithmetic tasks outlined here are designed based on an analysis of the state textbooks provided free of charge to students. A national-level document detailing specific learning objectives for each grade and subject has been prepared and released quite recently, towards the end of 2017. While the reading and arithmetic tasks are administered every year, ASER also tests some additional competencies. In previous years these have included basic English, applied arithmetic and reading comprehension, among others.

Also evident was the fact that, once children had fallen behind, the opportunities to acquire the abilities expected in the early years of primary school were scarce. Even in Class 8, close to one-fifth of all children were still unable to read at Class 2 level. As in the case of Class 5, this fraction decreased further after 2010. In 2016, the latest year for which ASER data are available, more than one-quarter of all students enrolled in Class 8 were unable to read a Class 2 level text. In other words, about one in four children is completing the eight years of free and compulsory schooling mandated by the Government of India without acquiring even foundational reading skills. It is apparent that the school system has been unable to cater to the learning needs of a student population that has expanded enormously in terms of both size and diversity in the space of just a few years.

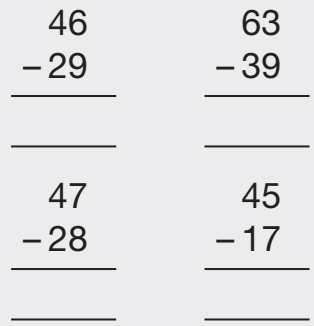
Turning to basic arithmetic abilities, the picture is similar, as illustrated in **Table 3.7**. In the years leading up to 2010, about seven in ten students in Class 5 could solve a two-digit numerical subtraction problem with borrowing, typically taught in Class 2 in Indian schools (an example can be seen in **Figure 3.23**). By 2016, only one-half of Class 5 students could solve a problem of this kind. As in the case of reading, even in Class 8 significant proportions of children had not mastered these basic arithmetic skills, and this proportion further declined after 2010.

**Table 3.7 Percentage of children from Classes 3, 5 and 8 who can do a Class 2 level subtraction (two-digit subtraction with borrowing)**

| Year | Class 3 (%) | Class 5 (%) | Class 8 (%) |
|---|---|---|---|
| 2008 | 38.8 | 69.8 | 88.5 |
| 2010 | 36.3 | 70.9 | 88.8 |
| 2012 | 26.3 | 53.5 | 73.7 |
| 2014 | 25.3 | 50.5 | 67.3 |
| 2016 | 27.6 | 50.5 | 66.5 |

*Source: ASER India.*

**Figure 3.23 Subtraction problems from the ASER tool, typically taught in Class 2 in Indian schools**




*Source: ASER India.*

Although not directly comparable with the ASER estimates, learning achievement data produced by the Government of India also point to declining learning outcomes among India's elementary school students. Aggregate national results from the latest round of the National Achievement Survey (NAS)[44], 23 of which were conducted in November 2017, are still awaited. However, the previous cycle of NAS for Class 5, conducted in 2014, concluded that the average achievement of Class 5 students on reading comprehension tasks declined from 2010 to 2014, as did the achievement of both the top 25% and the bottom 25% of students. In mathematics, a decline in average achievement from 2010 to 2014 was observed in every content area assessed (NCERT, 2015).

Poor and declining learning levels are also reported in other research studies. For example, the Young Lives study in the Indian state of Andhra Pradesh tracks cohorts of children over time. It concluded that a "comparison of scores in mathematics tests shows that learning levels have declined by 14 percentage points for 12-year-olds in 2013 compared with children of the same age in 2006" (Young Lives, 2014).

44 Designed by India's National Council for Educational Research and Training, NAS is a pen-and-paper assessment administered periodically to a sample of students in Classes 3, 5, 8 and 10 in government and government-aided schools that assesses student performance relative to grade level expectations. Different cycles of NAS have employed different methodologies for sampling and data analysis as well as different assessment instruments, making comparisons over time infeasible.

A second broad trend observable in ASER data is that, although children do acquire foundational skills as they continue in school and proceed to higher grades, the learning trajectories of successive cohorts are quite similar and low, as shown in **Figure 3.24**.[45] If a goal of the school system is to ensure that most children reach the learning outcomes expected of them at their grade level, then the learning curve for basic reading – a fundamental building block for all future progress in school – needs to be much steeper during their primary school years.

The third broad trend observable in ASER data is that each successive cohort seems to do worse than the previous one. For example, **Table 3.8** presents learning outcomes in arithmetic of three cohorts over time. Of the first cohort – those who were in Class 5 in 2007 – 42% could do division in Class 5 in 2007, as compared to 38% of the cohort that was in Class 5 in 2009. Of the children who entered Class 5 in 2011, only 28% could solve a similar division problem.[46]

**Figure 3.24 Percentage of children who can read a Class 2 level text**


*Source: ASER India.*

**Table 3.8 Percentage of children who can do division**

| | Cohort 1 in Class 5 (2007) | Cohort 2 in Class 5 (2009) | Cohort 3 in Class 5 (2011) |
|---|---|---|---|
| Class 5 | 42.4 | 38.0 | 27.6 |
| Class 6 | 50.0 | 50.1 | 33.1 |
| Class 7 | 59.7 | 48.3 | 38.8 |
| Class 8 | 68.3 | 48.0 | 44.1 |

*Source: ASER India.*

Each column in Table 3.8 can be seen as the learning trajectory for a specific cohort of students. It is clear from the data that in each class learning levels of each successive cohort are worse as compared to the previous cohort. For instance, by the time the 2007 Class 5 cohort reached Class 8 in 2010, 68% could do division. In contrast, only 48% of the cohort that started Class 5 in 2009 could do division by the time they reached Class 8 in 2012; and just 44% of the cohort that started Class 5 in 2011 could do division when they reached Class 8 in 2014. In other words, the learning trajectories of successive cohorts lie below those of previous cohorts *(see **Figure 3.25**)*. What this means is that each additional year of schooling is adding less for each successive cohort.

## 3.5.3 Conclusions

Several key challenges surface repeatedly from the evidence discussed. First, a substantial proportion

45 Ideally, to measure change in learning outcomes, the same children would be tracked over time. While ASER does not track children longitudinally, it can be used to create artificial cohorts to see how successive cohorts are faring as they move through different grades.

46 This analysis is for all children currently enrolled in school, whether government or private school. A similar analysis for only government schools shows that learning levels are lower as compared to private schools. However, it is well known that the demographic and background characteristics of private school children can be quite different from those of government school children – these need to be controlled for when comparing learning outcomes. But even children in private schools are far from reaching grade level expectations.

**Figure 3.25 Cohorts over time: Percentage of children from Class 5 to Class 8 who can do division**


*Source: ASER India.*

of students in India complete the eight years of compulsory schooling without acquiring basic literacy or numeracy skills. Second, when students do not acquire the capabilities expected of them in early primary grades, it is difficult to catch up in later years. As school enrolment expands to previously-unreached populations, many children in elementary school are first-generation school-goers, meaning that supplemental help at home is often not available. At the same time, much of the teaching in Indian classrooms focuses on transmitting the content of textbooks for that grade and targets children who are at grade level, with the result that those who have fallen behind do not get the opportunities or the support that would enable them to catch up. Being in this kind of "low learning trap" means that, although there is expenditure on schooling both by families and by the government for each year spent in school, the "value added" in terms of learning is minimal.

By including both enrolment and learning goals as part of SDG 4, the world now has a framework that acknowledges the fact that getting children to school is not enough. This is clearly reflected in Indicator 4.1.1. As shown in Table 3.6., ASER data from 2016 show that in Class 3 just one in four children can read a Class 2 level text; and even in Class 8 – the end of the elementary cycle in India – more than one-quarter of students are still unable to do so. This means that each year an estimated 6 million children complete elementary school in India but without having acquired even the basic skills required for future progress, whether academic or professional. The gap between rising expectations and falling ability levels poses a serious obstacle to India's ability to realise the promise of a "demographic dividend" due to its young population.

Today, ASER estimates are routinely quoted by those thinking about the quality of education in India. But these issues are not unique to India. The ability of the ASER assessment model to diagnose the core issues at the heart of the learning crisis using metrics and measures that are simple, quick, scalable, easy to understand and above all actionable has generated a ripple effect that has spread from country to country, leading to a unique South-South collaboration that is known today as the PAL Network.

Currently comprising 14 countries across three continents, each network member implements a citizen-led assessment (CLA) that follows a set of principles that is common across the network, adapting the tools and methods to the specific context of their own country. These principles include, for example, doing a household rather than school-based assessment in order to include all children; focusing on foundational reading and arithmetic abilities; and involving "ordinary citizens", among others.

The ASER tool is also at the heart of Pratham's model for remedial teaching, which is known as Teaching at the Right Level (TaRL), which uses the ASER assessment tool as a means to understand what children can do and then teaching them using

methods and materials designed to help them get to the next level. The TaRL model has been rigorously evaluated and found to be a highly effective means of improving children's foundational abilities. Much like the ASER model before it, its simplicity and scalability is finding uptake in many countries. Given the scale of the learning crisis worldwide, not only for children out of school but also for those already in the system, there is an urgent need to generate robust evidence that can be directly linked to action on the ground to improve learning outcomes.

## 3.6 THE ROLE OF TWAWEZA EAST AFRICA (UWEZO) CITIZEN-LED ASSESSMENTS IN TRACKING LEARNING OUTCOMES IN EAST AFRICA

The implementation of policies and strategies geared toward achieving the MDGs in education in the 2000's has led to huge progress in achieving universal primary education. By 2015, school enrolment rates in most developing countries increased to 95%. However, little was achieved in improving the quality of education. In order to propel a country to achieve its national goals, provision of quality education should go beyond access and adopt a system that develops knowledge, skills, values and attitudes. Thus, Education for Sustainable Development (ESD), which is enshrined within SDG 4, aims to ensure inclusive and equitable quality education that promotes lifelong learning opportunities to equip learners with relevant skills to tackle today's global, environmental and social challenges.

Uwezo is an example of a CLA that offers a platform to track learning outcomes in basic literacy and numeracy for Indicator 4.1.1 in Grade 2 or 3.

### 3.6.1 Uwezo and other CLAs[47]

Uwezo is conducted nationally at the household level in East Africa (Kenya, Uganda and the United Republic of Tanzania). The Uwezo message – and all CLA initiatives united under the PAL Network – that "Schooling isn't leading to learning" has gained traction globally. In late September 2017, the World Development Report 2018, *Learning to Realize the Promise of Education,*[48]was published. Its first main message stated that "schooling is not the same as learning". This was a core message that Uwezo has helped to reveal and amplify since 2009. It was inspired by India's ASER *(see Section 3.5)* and amplified by the PAL Network of CLAs.

Over the past decade, the growing family of household-based, citizen-led basic assessments of reading and arithmetic has proven that it is possible to engage citizens to measure basic learning outcomes of children and to use those results to spark change. In recent years, this innovative approach to learning assessment has been implemented in several Asian and African countries. Using basic reading and arithmetic tasks, organized groups of citizens in these countries have been systematically assessing for themselves what their children are able to do.

East Africa's Uwezo CLA initiative has several key features, common to all CLAs under the PAL Network. This is to ensure that all children are represented in the sample. SDG 4 is about education for *all children* but not all children are enrolled in school. Furthermore, daily attendance in school may be very low in some countries and therefore the household is the place to find most of the children.

CLAs use rigorous sampling methodologies to generate representative samples of children at national and sub-national levels. This unique feature of targeting all children enables Uwezo citizen-led surveys to provide better coverage of the target population. This is specially the case in hard-to-reach poorer areas that may be excluded from the international standardised school-based or household surveys. These surveys are the basis for many of the estimates used in assessing progress towards SDG 4 (Carr-Hill, 2017).

47 Written by James Ciera, Senior Data Analyst, Twaweza East Africa, Sara Ruto, Director, PAL Network, and Mary Goretti Nakabugo, Twaweza Country Lead and Regional Manager, Uwezo East Africa.

48 http://www.worldbank.org/en/publication/wdr2018

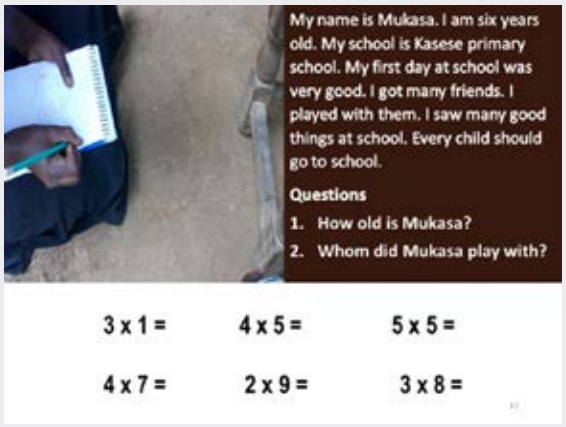
Second, the tools are designed to be simple so that parents, teachers, communities and ordinary people can conduct the assessment themselves and understand the findings. The simplicity of tools *(see example in **Figure 3.26**)* combined with the robustness of the results make this approach a powerful tool for change.

Third, when we assert that a certain percentage of children in a certain grade cannot read or do multiplication as laid out in Figure 3.26, everyone can understand what that means, whether in the village or in the national government. This helps to build public opinion. Participation by a wide cross-section of society helps enormously not only to bring the issue of learning to the centre of discussions of educational policy and practice in each country but also to create energy and urgency for immediate action.

In the literacy tasks, children were asked to read a letter or identify letter sounds from the alphabet, read a word, read a paragraph, and read a short story and answer two comprehension questions. The assessment placed children on one of five levels, ranging from "non-reader" to being able to read and comprehend a short story. The tasks were given in the assumed order of difficulty, starting from the simplest (letter identification), and those unable to perform a task were placed at the previous level in the sequence and not assessed further. For a child to be considered competent in literacy, they had to demonstrate the ability to read a story.

In the numeracy tasks, children were asked to recognise numbers and perform basic operations of addition, subtraction, multiplication and division. In Kenya and Uganda, the Grade 2 curriculum includes division, therefore children in both countries were asked to solve a division problem. In the United Republic of Tanzania, for a child to be considered competent in numeracy, he/she needed to solve a multiplication problem. Similar to the literacy assessment, the numeracy tasks were given in the assumed order of difficulty, starting with the simplest level (number recognition), and those unable to perform a task were placed at the previous level in the sequence and not assessed further. Those who were assessed on multiplication or division had already performed the addition and subtraction (United Republic of Tanzania) and multiplication (Kenya and Uganda) tasks successfully. Successful performance in multiplication (United Republic of Tanzania) and division (Kenya and Uganda) was treated as the indicator of full numeracy competency.

**Figure 3.26 Example of Uwezo reading and arithmetic tasks**




*Source: Uwezo.*

The comparability of performance across countries was based on the percentage of children (within the 6- to 16-year-old age group) reaching the highest level, i.e. ability to read a story for literacy and ability to multiply (numeracy). Multiplication was used to compare performance across the three countries because it was the highest level that included all three countries.

### 3.6.2 Tracking learning and inequalities

The Uwezo assessment can be used to track learning levels and to uncover inequalities in learning outcomes to inform progress towards the attainment of SDG 4 by all children. To illustrate this, we use data from Uwezo learning surveys, conducted in 2015 in 153 districts in Kenya, 159 districts in the United

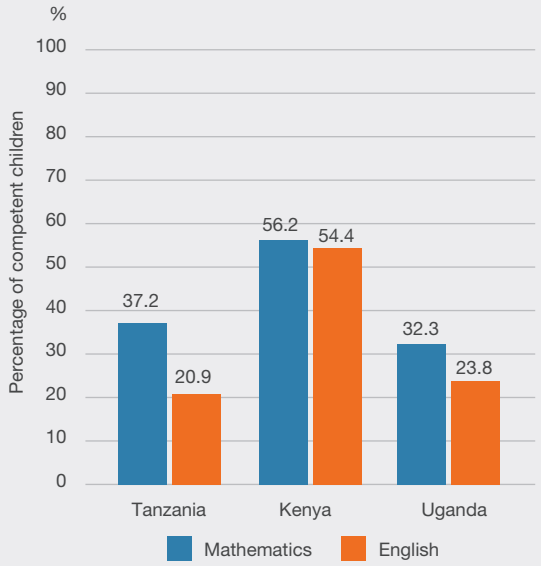
Republic of Tanzania and 112 districts in Uganda. The survey was administered to a nationally-representative random sample of children and youth within the 6- to 16-year-old age group. A total of 112,480 Kenyan, 104,267 Tanzanian and 94,248 Ugandan children and youth were assessed on competencies in literacy (up to story level) and numeracy (up to multiplication/division). The tasks were set according to the primary Level 2 curriculum in each of the three countries - the level to be attained after two years of primary education (aged 7 to 8 years). **Figure 3.27** presents the percentage of children able to do primary Level 2 numeracy (multiplication) and primary Level 2 literacy (reading a short English story).

**Figure 3.28** presents competence inequalities in both English and mathematics based on four social demographic characteristics:

- Mother's education: finished primary education or less versus some secondary education or more
- Household wealth: poor versus rich
- School type: public versus private[49]
- Sex: boy or girl

The figure shows the differences in the percentage of children/youth who attain the expected level of performance in numeracy and literacy, as a function of their demographic characteristics. The results indicate non-significant differences in the percentage of boys and girls reaching the expected performance level in the three countries. Private schools have slightly better learning outcomes compared to public schools. Maternal education and household wealth are the most deterministic factors that promote inequalities in learning outcomes. In Tanzania, the proportion of children/youth who reach the expected performance level is around 30 percentage points higher among those whose mothers completed at least some secondary education, as compared to the children/youth whose mothers at most finished primary education.

**Figure 3.27 Percentage of children (aged 6 to 16 years) competent in numeracy (mathematics) and literacy (English)**



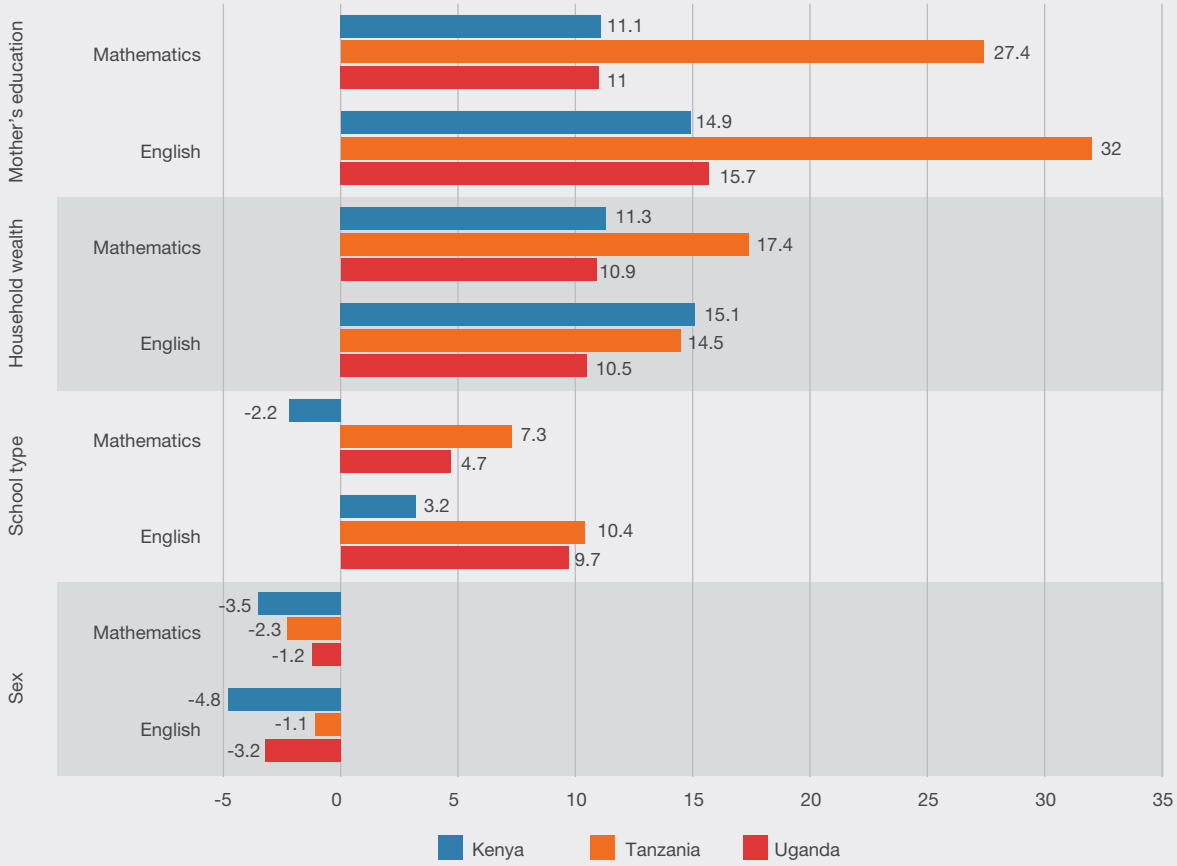

*Source: Uwezo.*

## 3.6.3 Conclusion

As international and national goals have moved beyond a focus on universal enrolment to universal learning, efforts like Uwezo and other CLAs can play a tremendous role in helping to track progress and identify problems..

49 The results for this category are based on the sub-sample of children/youth who are enrolled in schools.

**Figure 3.28 Learning inequalities: Differences in the percentage of children/youth reaching the expected performance level as a function of socio-demographic characteristics**




*Source: Uwezo.*

# 4. Reporting early childhood development

Target 4.2 focuses on early childhood development (ECD), care and pre-primary education in terms of quality and participation. It therefore presents a good example of a target that can be measured using administrative data and other sources of information.

The current global indicator for this goal is the "percentage of children under 5 years of age who are developmentally on track in health, learning and psychosocial well-being". Key concepts to measure include the quality of care and education, access to programmes, and child development and learning at the start of school. Measuring early childhood development is complicated but possible with sufficient technical consultation and operational support to countries in order to generate reliable data.

The idea of using one globally-comparable approach to measure ECD in all countries, rather than focusing on a region or group of countries (such as high- or low-income), is new. It is nonetheless informed by a long history of ECD measurement. The literature shows that for decades researchers and clinicians in a range of countries have developed and been using measures of ECD based on psychometric properties. Typically, these standardised scales have been tied to norms for use in high-income countries.

This chapter begins by discussing the challenges in measuring Indicator 4.2.1. Section 4.2 presents the vision of the global custodian agency, while Section 4.3 discusses a holistic view from an ECD expert.

## 4.1 HOW HAS ECD BEEN MEASURED TO DATE?[50]

In recent years, attention has focused on development of regionally- or globally-comparable population-based measures of ECD. All the tools summarised in **Table 4.1** are designed to capture children's development in the late preschool years using a combination of mathematics, literacy, language, and social/emotional and motor development items. Several measures are used across more than one country and at the population level *(see **Figure 4.1**)*.

There are advantages and disadvantages for each type of tool. Direct assessment is sometimes considered to be the most objective way to capture information on children's development. In many cases, this type of assessment may not be feasible unless it is carried out within a household survey and may not be capable of capturing many aspects of social/emotional development. Parents may not be accurate in reporting on specific details of their children's development as direct observers, even though they have the most depth and breadth of knowledge and therefore offer different information from that captured by other forms of direct assessment. Teachers are good reporters of children's behaviour in schools and therefore may be well-suited to predict which children will succeed over time, but only if they have the chance to get to know each child individually.

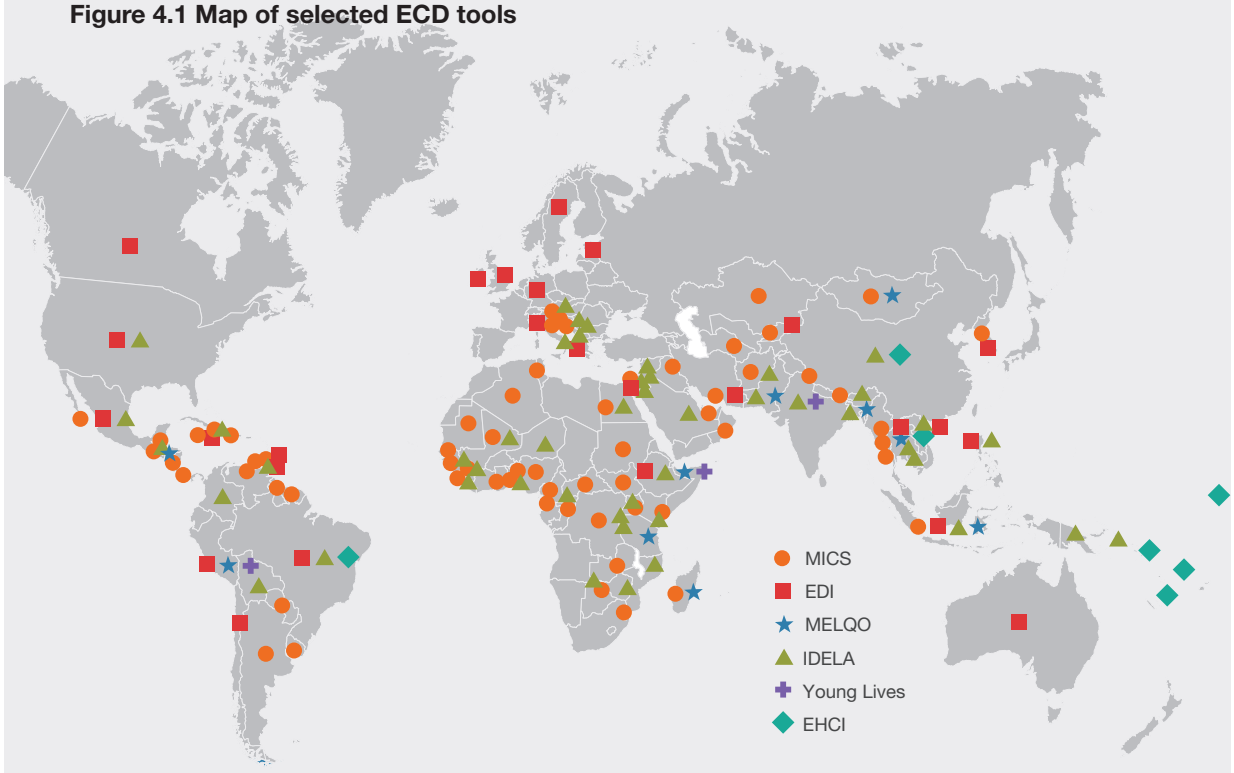
50 This discussion reflects the papers written by Anderson and Raikes (2017) and Yoshikawa et al. (2017) for GAML Task Force 4.

**Table 4.1 ECD measurement tools for 5-year-olds which have been tested in more than one country**

| Tool | Type of administration | Income level of country where it was tested |
|---|---|---|
| **East Asia Pacific Child Development Scales (EAP-CDS)** | Direct assessment | Middle-income |
| **Early Development Instrument** | Teacher survey | High-and middle-income income |
| **Early Human Capability Index (EHCI)** | Direct child assessment | Middle-income |
| **International Development and Early Learning Assessment (IDELA)** | Direct child assessment and caregiver survey | Low- and middle-income |
| **MICS Early Child Development Index (ECDI)** | Parent survey | Low- and middle-income income |
| **MELQO Measure of Development and Early Learning (MODEL)** | Direct assessment or parent or caregiver survey | Low- and middle-income |
| **Regional Project on Child Development Indicators (PRIDI)** | Direct child assessment | Middle-income |
| **UNICEF WCARO Early Learning Assessment (ELA) of Primary Education Entrants** | Direct child assessment and group assessment | Low- and middle-income |
| **Strengths and Difficulties Questionnaire** | Parent survey | Low-, middle- and high-income |

*Source: Anderson and Raikes, 2017.*

**Figure 4.1 Map of selected ECD tools**




Note: The ECD initiatives on the map are not indicative of any national or regional sampling.
*Source: UNESCO Institute for Statistics (UIS).*

### 4.1.1 Defining "globally-comparable"

The SDG target identifies health, learning and psychosocial well-being as key domains in determining readiness for primary school. Within each of these broad domains, a smaller sub-set of domains can be selected for global monitoring based on feasibility and desirability. There are a few key considerations to keep in mind when evaluating the extent to which the domains may be considered globally-comparable.

First, developmental science points strongly towards a holistic view of early childhood development, because early development is interconnected, with many skills supporting development across domains. This means that multiple domains are necessary to describe children's learning and development, regardless of comparability.

Second, some domains are more easily indexed than others. For some domains, internationally-comparable data may be easier to reliably achieve across countries because children typically follow a predictable pattern of progressively more complex development. However, the harder-to-index skills may be some of the most critical to measure (i.e. social/emotional development) and the least comparable across contexts.

Third, nearly all major assessments of child development include multiple domains, with different names and/or the same items, but often assigned to different domains. Assessment of comparability thus should include careful examination of constructs and items, as well as domains.

Finally, there is currently no systematic approach to determining standards for testing international comparability in early childhood. While there are certainly standards that can be applied from primary school learning measurement, the unique nature of ECD means that a specific set of standards should be developed and applied, before determining whether existing data point towards comparability or lack thereof in domains.

### 4.1.2 "Developmentally on track"

There are presently no agreed-upon definitions of "developmentally on track" that are specific enough to guide internationally-comparable, regional or national measurement. Conceptually, identifying some children as "developmentally on track" implies that other children are not "developmentally on track" simply by the nature of the statement, which is articulated as a binary option (either "on" or "off" track).

**Box 4.1. Issues in globally-comparable measurement**

An immediate step is to decide on standards for international comparability in early childhood data and to assess existing data sources against these standards.

There are potential tensions between feasibility and precision, and the challenges are technical and with financial constraints.

Household surveys are typically more expensive than centre- or school-based assessments, because it is necessary to sample and visit individual households. Less travel time is required when a group of children is in one location.

Theoretically, all domains of child development could be measured in an internationally-comparable way. By considering existing data and finding a balance between feasibility and desirability, the GAML Task Force on Indicator 4.2.1 discussed the possibility of a stepping-stone strategy (e.g. starting with the easiest domain or what would be measured in subsequent levels of education), but this was discarded (see discussion below for more information).

*Source: UNESCO Institute for Statistics (UIS).*

**Table 4.2 Options for defining "developmentally on track"**

| Method of comparison | National standards | Creation of global scale | Leave undefined |
|---|---|---|---|
| **Absolute** | Percentage of children reaching agreed-upon set of skills/competencies, using national standards as starting point. | Set of skills defined by experts but no "absolute" threshold because would be structured as a relative scale. | Up to countries to define standard set of skills to measure against, could look across countries over time to identify points in common. |

*Source: Anderson and Raikes, 2017.*

**Option 1: Rely on national standards.** Many countries have gone through the process of developing early learning development standards (ELDS) or other types of standards that include children's development. These standards are holistic in nature and are intended to inform measurement by outlining consensus on what children should be able to do at certain ages. This approach runs the risk of perpetuating inequity because the quality of the standards and the extent to which the standards are developmentally appropriate may vary considerably by country. To generate a globally-comparable estimate, the purchasing power parity (PPP) estimate is calculated, based on a common global currency scale. The applicability of this approach to ECD could be explored as a path towards synchronising national-level and globally-relevant data.

**Option 2: Invest in the creation of a global scale.** The World Health Organization (WHO) invested in the development of growth scales that have had a profound impact on attention to malnutrition. A similar approach could be explored for older children as well. A first step would be careful examination of the pros and cons of the feasibility and desirability of this approach, including costs and expected benefits.

**Option 3: Leave undefined.** Assume that "developmentally on track" is useful as a conceptual model but that it cannot be precisely quantified and therefore will not be measured anytime soon. Overtime, it could be informed empirically by using existing data to more fully define a cross-nationally-relevant definition.

GAML Task Force members proposed a hybrid approach between Options 1 and 2, where national standards are reviewed and used to develop a global definition of "developmentally on track" and a possible global scale.

## 4.2 MONITORING EARLY CHILDHOOD DEVELOPMENT OUTCOMES IN THE SDGS[51]

ECD is a maturational and interactive process involving an ordered progression of motor, cognitive, language, socio-emotional and regulatory skills and capacities across the first few years of life. During these years, a child's newly-developing brain is highly plastic and responsive to change as evidenced by the billions of integrated neural circuits that are established through the interaction of genetics, environment and experience. This makes early childhood a critical time for cognitive, social, emotional and physical development and sets the stage for lifelong thriving.

The importance of ECD as a necessary and central component of global and national development has been recognised by the international community through the inclusion of a dedicated target and indicator within the SDGs. Target 4.2 specifically calls upon countries to "ensure that, by 2030, all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education". One of

51 Written by Claudia Cappa and Nicole Petrowski, UNICEF. UNICEF is the custodian agency for Indicator 4.2.1.

the indicators selected to measure this target is Indicator 4.2.1 (the percentage of children under 5 years who are developmentally on track in health, learning and psychosocial well-being). ECD is also linked to the achievement of other SDG targets, including those related to eradicating poverty and hunger, promoting economic growth and productivity, attaining gender equality, and building peaceful and inclusive societies.

### 4.2.1 Measuring ECD in household surveys

Measuring children's development is a complex undertaking. While the overall developmental process is similar across cultures, children develop at different speeds and may reach developmental milestones at different times. What is considered "normal" child development also varies across cultures and environments, since expectations and parenting strategies may differ not only among countries but also among cultural, ethnic or religious groups within the same country. Finally, child development encompasses many dimensions of wellbeing, all of which need to be measured to provide a comprehensive assessment of children's development outcomes and possible risk factors.

UNICEF has been working with countries to collect data on ECD through the Multiple Indicator Cluster Surveys (MICS), a global household survey programme that produces statistically-sound, nationally-representative and comparable data on several key indicators of the health and wellbeing of children, women, men and families.

MICS questionnaires cover several aspects of child development and wellbeing, including access to early childhood care and education, nutritional status, immunisation and parenting practices, as well as the conditions and quality of care within a child's home environment, for example, the availability and variety of learning materials in the home, early stimulation and responsive care and non-adult supervision.[52]

In order to capture information on children's achievement of universal developmental milestones across countries, UNICEF formed a technical advisory group in 2007 to develop, within the context of MICS, a set of specific questions posed to mothers/caregivers to measure the overall developmental status of children within physical, literacy-numeracy, social-emotional and learning domains. Following a review of existing tools, consultations among a broad group of experts, and field-testing and validation, a 10-item index – ECDI – was added to MICS beginning with the fourth round of surveys, primarily implemented between 2009 and 2012 *(see **Figure 4.2**).*[53]

The ECD data from MICS have been used in a number of academic articles and data-driven advocacy flagship reports (UNICEF, 2018; Miller et.al., 2016; McCoy et.al., 2016; Jeong, Bhatia and Fink, 2018) and data quality has been analysed through various reliability and validity tests (Kariger et al., 2012). For instance, the validity of the ECDI was confirmed through an analysis of data collected in 12 countries during the fourth round of MICS and a number of studies have also conducted cross-country comparisons using the index (McCoy et.al., 2016b; Miller et.al., 2016). With the inclusion of the ECDI, MICS has become the largest source of comparable data on developmental outcomes for children, producing country-level estimates for more than 60 mostly low- and middle-income countries.[54]

When UNICEF started the process of creating a tool for measuring ECD outcomes in household surveys,

52 Learning materials include books and play materials, which are defined as household objects, objects found outside (such as sticks, rocks, shells, etc.), homemade toys and manufactured toys. Activities that provide early stimulation and responsive care include: reading books to the child; telling stories to the child; singing songs to the child; taking the child outside the home; playing with the child; and naming, counting or drawing things with the child.

53 The literacy-numeracy domain is captured by ECDI Items EC6, EC7 and EC8, while the learning domain is measured by ECDI Items EC11 and EC12.

54 Some countries collected data on ECDI in multiple rounds of MICS. The ECDI has also been collected in approximately ten countries through its inclusion in demographic and health surveys.

**Figure 4.2 UNICEF's Early Childhood Development Index (ECDI)**

| |
|---|
| **EC6**. I would like to ask you some questions about the health and development of (**name**). Children do not all develop and learn at the same rate. For example, some walk earlier than others. These questions are related to several aspects of (**name**)'s development.<br><br>Can (**name**) identify or name at least ten letters of the alphabet? |
| **EC7**. Can (**name**) read at least four simple, popular words? |
| **EC8**. Does (**name**) know the name and recognise the symbol of all numbers from 1 to 10? |
| **EC9**. Can (**name**) pick up a small object with two fingers, like a stick or a rock from the ground? |
| **EC10**. Is (**name**) sometimes too sick to play? |
| **EC11**. Does (**name**) follow simple directions on how to do something correctly? |
| **EC12**. When given something to do, is (**name**) able to do it independently? |
| **EC13**. Does (**name**) get along well with other children? |
| **EC14**. Does (**name**) kick, bite, or hit other children or adults? |
| **EC15**. Does (**name**) get distracted easily? |

**Note:** The response options for each questions are yes, no or don't know.
*Source: UNICEF.*

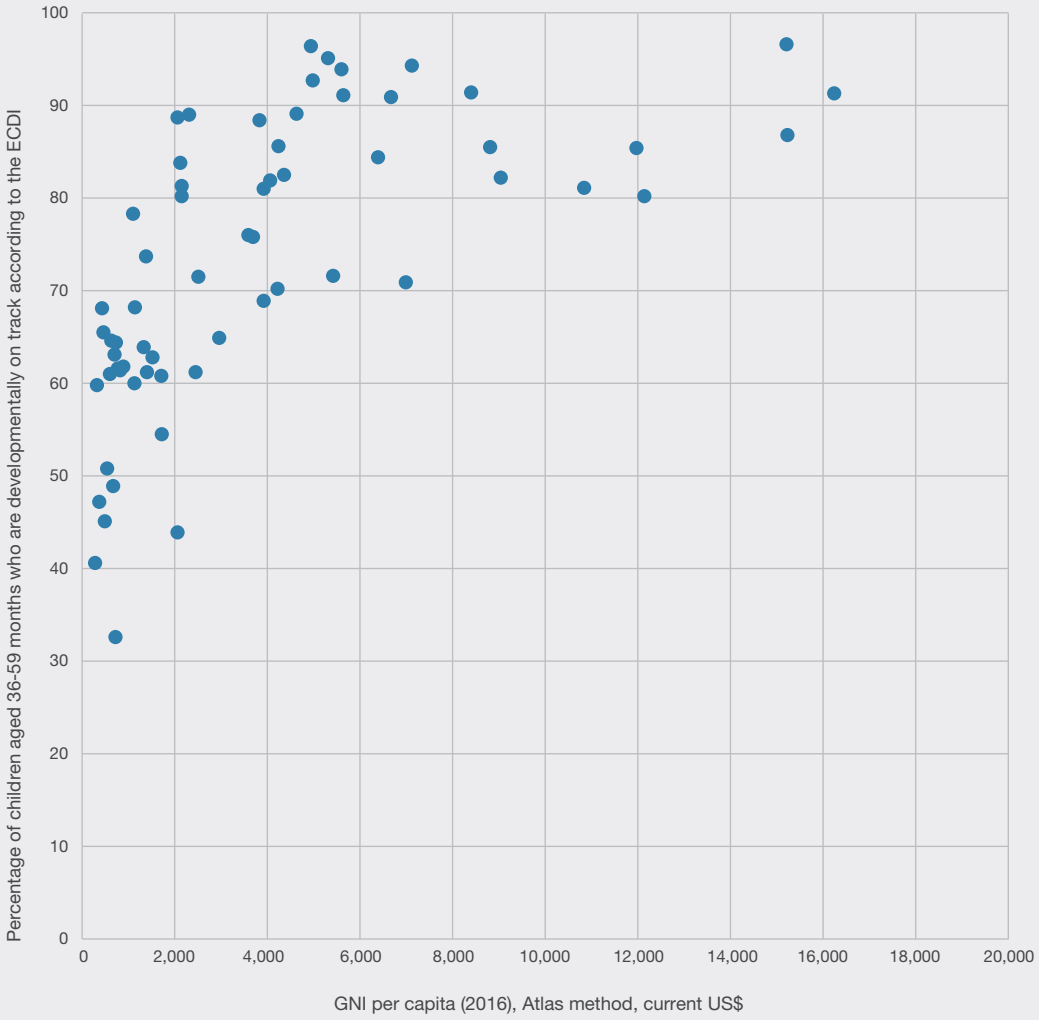
there was only a handful of available measures with the aim of collecting data on child development outcomes at the population level in order to produce representative national prevalence estimates as opposed to evaluating interventions or conducting clinical assessments of individual children. However, the landscape has changed since that time, and a number of groups have been working to develop, test and validate measures of ECD with various purposes in mind. In many instances, these other tools rely on direct assessment of children and/or teachers' reports and are not designed to produce representative estimates at the national level. Furthermore, only a few of these tools have been tested or used to collect data across a variety of country contexts, and some of the available measures have only been designed and validated for use in certain countries or regions. This has limited the ability to make cross-country comparisons or to reliably aggregate data into global and regional estimates of child development.

### 4.2.2 Evidence on child development outcomes collected through the ECDI

In 68 countries with comparable data generated through the implementation of the ECDI for the period 2010 to 2017, around two in three children aged 3 to 4 were developmentally on track in at least three of the following domains: literacy-numeracy, physical development, social-emotional development and learning.[55] In all countries with available data, more than 80% of children between the ages of 3 and 4 are considered to be on track in their physical development. With regard to learning

55 The four domains are defined as follows:
Literacy-numeracy: Children are identified as being developmentally on track if they can do at least two of the following: identify/name at least ten letters of the alphabet; read at least four simple, popular words; and/or know the name and recognise the symbols of all numbers from one to ten.
Physical: If the child can pick up a small object with two fingers, like a stick or rock from the ground, and/or the mother/primary caregiver does not indicate that the child is sometimes too sick to play, then the child is regarded as being developmentally on track in the physical domain.
Social-emotional: The child is considered developmentally on track if two of the following are true: The child gets along well with other children; the child does not kick, bite or hit other children; and the child does not get distracted easily.
Learning: If the child follows simple directions on how to do something correctly and/or when given something to do, and is able to do it independently, then the child is considered to be developmentally on track in the learning domain.

**Figure 4.3 Percentage of children aged 36 to 59 months who are developmentally on track in at least three of four domains of child development (as measured by the ECDI) and gross national income (GNI) per capita in 2016 according to the Atlas method in US$, in countries with available data**




**Notes:** Each dot represents a country. Only those countries with data on both the ECDI and GNI per capita are included in this chart.
*Source: Data on GNI per capita are from World Bank national accounts data and OECD national accounts data files available at <https://data.worldbank.org/indicator/NY.GNP.PCAP.CD>. Data on the ECDI are from UNICEF global databases, 2018, based on MICS and other nationally-representative household surveys, 2010–2017.*

and social-emotional development, the proportions of children on track vary widely across countries but are above 50% in practically all countries with data. Children are least likely to be considered developmentally on track in the area of literacy-numeracy across all countries.[56]

**Figure 4.3** shows the relationship between the prevalence of children who are developmentally on track and national income per capita. This seems to indicate that most high- and upper-middle-income countries with available data generally have a relatively high proportion of children aged 3 to 4 considered to be developmentally on track with few exceptions.[57] On the other hand, there are noticeable disparities among children living in low- and lower-middle-income countries, with wide differences in the proportion of children developmentally on track even in some countries with similar income levels.

Access to high-quality care and education programmes outside the home can provide children with opportunities to develop the basic cognitive and language skills they need to flourish, build social competency and foster emotional development. Across all countries with data, children who attend early childhood education are found to be around two times more likely, on average, to be developmentally on track in the literacy-numeracy domain compared to children not attending early childhood education programmes.[58] Despite its proven benefits and clear impacts on children's early learning, nearly 57 million children aged 3 to 4 (just over two in three) do not attend an early childhood education programme in the 67 mostly low- and middle-income countries with available data.

### 4.2.3 The need for an improved measure of ECD to monitor SDG Target 4.2

Currently, Indicator 4.2.1 has been classified as Tier III, meaning that the Inter-Agency and Expert Group on SDG Indicators (IAEG-SDG) has decided that methodologies and standards for measurement do not currently exist and need to be developed and tested. As the custodian agency of this indicator, UNICEF has been tasked with the responsibility of undertaking methodological work to develop, test and validate a survey module that can be used to collect nationally-representative data using a standardised approach and measure in order to monitor and track progress towards achieving Target 4.2. In the interim, the ECDI is being used as a proxy measure to report on Indicator 4.2.1, and for the past three years, ECDI data have been featured in the United Nations Secretary-General's report, *Progress towards the Sustainable Development Goals,* and the accompanying statistical annex.

There are several key reasons that necessitate the development of an improved measure of ECD within the context of SDG monitoring and reporting. Currently, the main differences between the existing ECDI and the formulation of Indicator 4.2.1 pertain to the inclusion of the health domain and the broader age group of children under 5 years in the SDG formulation. In addition, the principle of universality within the SDG agenda and the need to ensure that tools are relevant and applicable for all countries also needs to be considered. The intention is to build an improved measure of ECD that will be aligned with the definition set by Indicator 4.2.1. A comparative advantage of this measure is that it is being designed for integration into existing national data collection efforts and will not require the implementation of a separate, dedicated survey effort, which are often time- and resource-intensive. These population-level data, collected as a component of household surveys, will allow disaggregation of the findings by key demographic and socioeconomic characteristics, as well as sub-national geographical areas.

56 UNICEF analysis based on data from MICS and other nationally-representative household surveys, 2010–2017.
57 These results are partly skewed given the limited data availability for high-income countries.
58 UNICEF analysis based on data from MICS and other nationally-representative household surveys, 2010–2017.

The methodological work is being led by UNICEF, in collaboration with an expert advisory panel and under the auspices of a Global Inter-Agency Expert Group on ECD Measurement tasked with overseeing the revision, testing and validation of the improved measure of ECD outcomes.

Key activities completed to date include: a scoping exercise and review of more than 500 items that assess ECD through both caregiver and teacher reports, as well as direct assessments included in ten existing tools/instruments; cognitive testing of a bank of items in six countries (Bulgaria, India, Jamaica, Mexico, Uganda and the United States); and commissioning of a series of background papers on young children's development in health, learning and psychosocial wellbeing to inform the development of a conceptual framework and a report on psychometric considerations to ensure the development of a strong tool/instrument for measurement purposes. UNICEF has also hosted a series of technical consultations in 2015, 2016 and 2018 to bring together academics, technical experts and key partners in the field of ECD measurement and tool development in support of the methodological work.

A dedicated field test of the measure will take place in Mexico in 2018. Following this, additional testing, validation and piloting of the measure in a number of selected countries may follow. By the end of the process, the final output will be a standardised and validated tool to measure ECD outcomes, along with guidance on its implementation that can be broadly used by countries in national household surveys for monitoring of SDG Target 4.2. As is the case with all MICS tools, the improved measure will be a public good that will be freely accessible to all countries interested in undertaking data collection on ECD at the population level.

## 4.3 PATHS TO EQUITABLE MONITORING OF EARLY LEARNING WITH SDG 4[59]

The anthropological theory of culture (Goodenough, 1994), applied to children's cognitive and social learning (Cole and Cagigas, 2010; Goodnow, 1990, 2010), posits that learning is both a cognitive and a social process, leading to the creation of the knowledge that a child needs to successfully participate in its society. While the cognitive processes of learning may be universal, their results are interpreted and transformed through cultural practices, customary behaviours and ways of life. Applying this perspective to a global level of measuring of learning is challenging, yet crucial to achieve comparability and equitable progress.

Despite marked global progress in the enrolment of children in primary education over the last decade, many children still do not have access to education. This is especially the case for the youngest ones, those in the poorest households and in conflict areas. Spurred by this evidence, the efforts of the international community on early learning focused on improving the specificity of early education targets in the new SDGs. Target 4.2 aims to: "By 2030, ensure that all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education". This Target is accompanied by a set of indicators meant to inform the monitoring frameworks in tracking progress towards the achievement of the SDGs by 2030.

As previously explained, Indicator 4.2.1 is "the proportion of children under 5 years of age who are developmentally on track in health, learning and psychosocial wellbeing, by sex". It is the expectation of UN Member States that the global expertise in early education will contribute to the development of equitable measures of these indicators.

59 Written by Magdalena Janus, Offord Centre for Child Studies, Department of Psychiatry and Behavioural Neurosciences, McMaster University.

Reliable and equitable accountability mechanisms are needed to measure the progress towards, and eventual achievement of, the targets. Since their endorsement, the field of early education has repeatedly declared its commitment to contribute to the effort of keeping the SDGs on track (Raikes et al., 2017). Emerging voices are calling for scientific rigour and breadth in the measurement of the SDGs. Issues with external and internal validity, the incorporation of longitudinal studies and most of all a reliance on developmental science are among the factors being considered (Verma and Petersen, 2018). Inclusion of the goal focusing on early childhood education in the SDGs is an acknowledgment of the importance of early development and its contribution to, not only individual healthy trajectories, but the health of nations and prosperity of the world as well.

### 4.3.1 Opportunities and challenges

Focus on early development, embedded as it is in the context of other priorities, offers both opportunities and challenges. The target and indicator descriptors are, by the nature of the complex document they are part of, static, addressing a status that can be described by a number. It is up to the measurement, education and developmental science community to add dynamic character and depth to those indicators, by learning to use data to optimise trajectories of learning and developing innovative strategies to address the data gap in very early development.

One of the first opportunities offered by Target 4.2 is the motivation for broadening the scope of developmental and educational sciences beyond what has traditionally been used as gold standards. Target 4.2 and its indicators make it imperative to provide reliable tools and methodologies to learn about ECD, universalities and idiosyncrasies, taking into account cross-national, ethnic, geographic and (dis)ability boundaries while developing best practices, optimal outcomes and customised approaches. While some of these opportunities involve the creation of new tools (McCoy et.al., 2016), or overhauling existing ones like UNICEF's ECDI,[60] these initiatives should go hand in hand with: i) an innovative use of existing, historical data; ii) the creation of platforms for data-sharing and storage; iii) the development of techniques for data harmonisation; and iv) an expansion, rather than replacement, of support for locally-relevant data collection in order to broaden the scope and increase the contextual understanding of whether progress has been achieved. While the formulation of Target 4.2 is very specific and there should be one common way of addressing its measurement globally in the short term, its interpretation cannot, ultimately, be confined to one single number devoid of depth or context.

If the opportunities are beguiling, the challenges are equally daunting. Three in particular stand out and must be accounted for in the measurement framework.

First, while expedient and necessary in the short term, adopting one measurement as the only means to monitor learning and development is neither realistic nor necessary for making progress, or appropriate for understanding the course of change.

Second, the inherent sources of error in measurement are many. One of the most important and most obvious is the mode of administration of the assessment, which has to be adequate for the developmental age and appropriate for the assessment goal. It has to be acknowledged and agreed that no administration mode (direct observation, interview with a caregiver or teacher report) is either fully objective or fully comprehensive (in the sense of including multiple domains in depth), and therefore, an effort must be made to include different administration modes and informants. The collection of data and measurement practices have to be accompanied by resources for data analysis and refinement, with ongoing checks for estimating reliability.

60 See Chapter xxx "Monitoring early childhood development outcomes in the SDGs" for a discussion on UNICEF's ECDI tool.

Finally, the aspiration for breadth and depth in measurement notwithstanding, it will be a difficult task to ensure that most or all countries report on one – let alone more than one – type of measurement.

It is often said that one cannot reliably measure early development. If measured using an interview with a caregiver, the respondent might have difficulty answering questions. If measured through a direct assessment, the reliability and training of assessors may be limiting. If exclusively measured by a physical health indicator, such as stunting, the output becomes a dichotomous indicator of development delay rather than the broad probability of being on track for optimal trajectory of development (Black et al., 2016). Despite a high level of reliability evidenced by existing data, the major limitation of school-based teacher reports is the age of the students, as teachers can only serve as informants for children who attend centre-based learning (Janus and Reid-Westoby, 2016).

Despite these concerns, a recent analysis of data for children under 3 years old from a variety of countries, collected with various instruments (some locally-developed), demonstrated a remarkable stability in arriving at a comparable, normative curve of the child development trajectory, regardless of the country or tool the data came from (Lancaster et al., 2018).

### 4.3.2 A potential way forward

Bearing in mind these opportunities and challenges, what are the consequences for building a measurement framework that could be used to report on early childhood for the SDGs?

It is undisputable that the vision for the new set of measures has to be broad. One effort – the revised ECDI – is aiming for a short, globally-comparable metric that could be easily interpreted. That effort needs to be strongly supported and endorsed, but at the same time the uniform effort should not come at the cost of suppressing the diversity of measurement. One potential solution is that, once established, the ECDI should become a component of as many local learning/development measurement and evaluation initiatives as possible to ensure that Target 4.2 is monitored in a sensitive and feasible way.

Monitoring child development at the population level[61] aims to address and counter bias in sample selection and has been implemented successfully in several countries, such as Australia, Brazil, Canada, Jordan and Kyrgyzstan using the Early Development Instrument (EDI) (Janus and Reid-Westoby, 2016). Repeated implementations over time can assist with understanding the trends and account for changes from one time point to another at the regional level. The advantage of the EDI is its comprehensive coverage of developmental domains, which requires considerable time for translation, adaptation and completion, especially at the population level. Offering a smaller developmental coverage of over 50 countries are existing data for many low- and middle-income countries, collected through nationally-representative samples with UNICEF's ECDI through MICS[62], which broadly reflect the developmental status of young children. In between the two levels of coverage, there are many databases that provide information on the developmental status of young children and could be harnessed to inform the current level of child development, to report on "on track" or normative development as stated in Indicator 4.2.1, and provide a baseline for further monitoring of the progress towards achieving Target 4.2.

Global efforts to achieve uniformity have the advantage of optimising resources and using highly-skilled professionals to examine the conceptual and psychometric quality of assessments. They should draw on local and regional expertise. It is simultaneously *imperative not to suppress the existing diversity of measurement*. Monitoring progress towards the achievement of goals, such as the SDGs, has to come with the capacity of understanding the variations across groups of interests, such as

61 For example, including all or nearly all possible participants, similarly to a census approach.
62 MICS, housed and managed by UNICEF, is the largest source of statistically-sound and internationally-comparable data on women and children worldwide (http://mics.unicef.org).

countries or regions. That understanding can only come with a broader perspective on several fronts: coverage (which encompasses inclusion of children with disabilities), mode and ease of administration, cultural applicability and local relevance. It may also be the case that some countries simply prefer to report based on their own national efforts or a global one they have become accustomed to, for various reasons, even if these are, in the opinion of global psychometricians, less than perfect.

Moreover, the similarity of the performance of items and scales is an extremely useful feature, but it needs to be complemented with culturally-sensitive (rather than culturally-neutral) means of assessment, by collecting locally-relevant information that may be idiosyncratic for a country or a region. A body of work exists to demonstrate that different tests can be interpreted on comparable metrics (Kolen and Brennan, 2014), and more effort should be directed towards the extension of that methodology to existing (and new) learning measurement tools. The second implication for the measurement framework is the continuous monitoring of the comparability and reliability of collected data.

There are three activities that the international community should engage in to not only report on results of measurement but to interpret them as well. They include:

i) Promoting a short, feasible and "universal" assessment, its application with all children and its use in a longitudinal framework;
ii) Facilitating the continuous use of validated tools that address more comprehensive development, from a variety of perspectives, in a culturally- and disability-sensitive, rather than neutral, way; and
iii) Enabling the collection of contextual, socioeconomic, demographic, educational and health service data that could assist in interpreting Indicator 4.2.1.

This effort should also be supported by three methodological and training initiatives:

i) Investing in innovative, cost-effective, time-saving and customised modes of data collection (tablets, mobile phones, etc.);
ii) Promoting statistical expertise in the assessment of the quality and comparability of measurement, such as assessment of measurement invariance, and further developing methods for cross-comparability of data collected by different tools; and
iii) Promoting training and expertise in the adaptation and use of global tools, such as the understanding of and adherence to criteria for modifications at the local or national level.

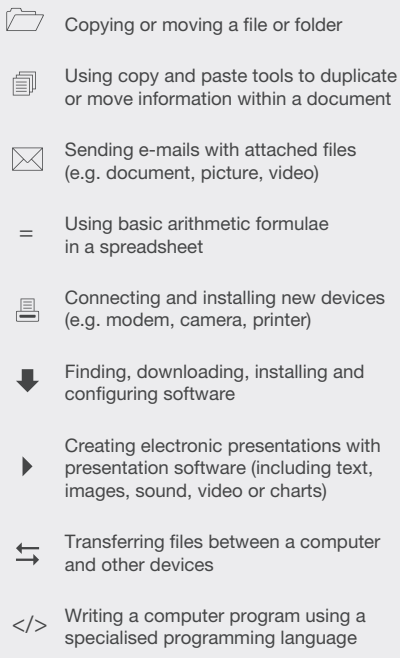
Reliable and relevant data collected in a culturally- and developmentally-appropriate manner are key to fulfilling the promises of the SDGs. They are needed by practitioners and policymakers to make informed decisions, by evaluating existing and new initiatives in a bias-free way that is translatable into action (Raikes, Dua and Britto, 2015; McCoy et.al., 2016). In addition, it is crucial that the results from data collection processes across countries and regions are comparable. The aim is not to create league tables but to highlight, understand and act on the progress (or lack of it) that may be determined by vastly differing contexts in diverse regions of the world. The opportunity which SDG 4 presents us with must be used to consider children in a global sense rather than splitting them into arbitrary categories.

# 5. Skills in a digital world

The current context of global development is characterised by an acceleration in the development, complexity and use of information and communication technology (ICT). Ensuring that everybody has access to ICTs is among the challenges (the first digital gap).

This chapter focuses on Target 4.4: "By 2030, substantially increase the number of youth and adults who have relevant skills, including technical and vocational skills, for employment, decent jobs and entrepreneurship". It explores Indicator 4.4.1: "the proportion of youth/adults with ICT skills, by type of skill". Both the target and the indicator reflect a forward-looking commitment by countries. But what does it mean to have such skills, and how can this be measured?

**Figure 5.1 Skills to be measured to assess ICT skills**




*Source: UNESCO Institute for Statistics (UIS).*

The global target concept for Indicator 4.4.1 argues that ICT skills determine the effective use of ICTs. The indicator is defined as the percentage of youth (people aged 15 to 24 years) and adults (aged 15 years and older) who have undertaken certain computer-related activities in a given period (e.g. the previous three months) *(see **Figure 5.1**).*

The global indicator is usually derived from a national ICT survey that typically asks a number of questions on access to various devices and the Internet within the household, and then asks one or more randomly-selected individuals from the household to answer questions on ICT usage, which includes skills. The indicator is calculated as the percentage of people in a given population who say "yes" when asked if they have used ICT skills, for example, inside or outside their school or workplace, have used those skills for a minimum amount of time, and have access to the Internet.

It is interpreted as the link between the use of ICT and its impact, which helps to measure and track the proficiency level of users. A high value indicates that a large share of the reference population has the ICT skill being measured.

Currently, there is one data source for this indicator based on the methodology adopted by the International Telecommunication Union (ITU). Eurostat collects the data annually for 32 European countries, while the ITU is responsible for setting up the standards and collecting this information from remaining countries.

One of the main measurement challenges for this indicator is the narrow coverage of "relevant skills" proposed by the target. In addition, it is based only on the information that people themselves report. They provide information on the types of activities they have undertaken but not their proficiency level.

It is impossible to verify the accuracy of their self-assessments, and more importantly, reporting varies markedly between groups from different cultural and personal backgrounds. Women, for example, tend to under-report their abilities in using computers and the Internet, while men tend to overstate them.

To develop a measurement strategy for Target 4.4, it is essential to address the following questions:

- What concept should be measured and how should it be defined? What do we mean by ICT skills or digital literacy? Should technical and vocational skills be considered as well?
- What measurement tool needs to be developed and how? Do we need different tools for different age groups (in particular for young people)?
- Should measures be equally appropriate for youth and adults in all countries, and if so, how can such scales be created?
- How will it be distributed to countries? How can countries be supported to implement the new tool?
- What is the cost of implementing the tool?
- How can we set baselines?
- With what frequency should countries measure and report?
- Consideration should also be given to the process of inserting the new indicator into the global list. Is this possible? If so, when and how?

This chapter describes the initiative, being led by the UIS, to develop thematic Indicator 4.4.1, as well as the experience of the European Union. Following the introduction, the discussion focuses on the work of the GAML Task Force 4.4. Section 5.3 describes the only existent cross-national framework for youth and adults. Section 5.4 presents the proposed Global Framework of Digital Literacy before Section 5.5 describes efforts underway to map existing tools to assess digital skills in youth and adults.

## 5.1 MEASURING DIGITAL LITERACY SKILLS: A MOVING TARGET[63]

SDG Target 4.4 focuses on a critical education outcome: skills for work. It is complementary to SDG Target 4.3, which refers to opportunities for technical and vocational education as a means of acquiring these skills. However, skills for work are acquired in all education programmes, not just technical and vocational ones. They can also be acquired outside formal systems of education and, instead, within families, communities and workplaces throughout the course of a lifetime.

From a global comparative perspective, it is not immediately clear what these skills are. Skill requirements are specific to jobs, which differ enormously across countries. Other than the foundational skills of literacy and numeracy, which are the focus of SDG Target 4.6, it is difficult to think of skills that satisfy three key criteria:

- Relevance in various labour market contexts;
- Can be acquired through education; and
- Measurable along a common scale at low cost.

The recommendation of the Technical Advisory Group (TAG) for SDG 4 indicators, which was adopted by the IAEG-SDGs, was to focus on ICT skills. While this narrowed the scope of the "skills for work" concept, it advanced the international education agenda, which until recently has ignored education outcome measures. Moreover, ICT skills meet the three criteria – they are increasingly relevant in diverse work environments around the world, can be taught in education programmes, and in theory are amenable to measurement.

In practice, concerns about cost-effective measurement led to the choice of Indicator 4.4.1: "the percentage of youth and adults with ICT skills by type of skill". According to the definition of the International Communications Union (ICU), data are collected

63 Written by Manos Antoninis, Director, Global Education Monitoring Report, UNESCO.

through household surveys or censuses and refer to nine computer-related activities that individuals report having undertaken in the previous three months. These range from copying and pasting, to using arithmetic formulae in spreadsheets, to writing computer programmes.

Despite being straightforward to interpret and collect, Indicator 4.4.1 reflects only the prevalence of certain computer-related activities and not the skill level at which they are performed. Such skills cannot be self-reported but need to be assessed directly. This led to the proposal of thematic Indicator 4.4.2: "the percentage of youth and adults who have achieved at least a minimum level of proficiency in digital literacy skills". The definition and development of this indicator is the focus of the GAML Task Force on Target 4.4.

Analysis for the 2017/2018 Global Education Monitoring (GEM) Report of 16 European countries, which collected data on skills indirectly through Eurostat household surveys and directly through the OECD Programme for the International Assessment of Adult Competencies (PIAAC) survey on problem-solving skills in technology-rich environments, showed that the two indicators (4.4.1 and 4.4.2) were correlated. This correlation was higher:

- In simple skills, e.g. sending emails with attachments, than in complex ones (e.g. programming); and
- At a lower level of PIAAC proficiency (i.e. Level 1, which corresponds to the use of widely-available applications to access information to solve a problem) than at a higher level (i.e. Level 2 and above, which requires the use of these applications to actually solve problems).

While the global indicator captures differences in ICT skill distribution among countries, it only does so at the most basic proficiency level (familiarity with applications). Countries are more interested in the acquisition of more sophisticated skills, which can make a difference in their economies.

### 5.1.1 Defining a framework of digital literacy skills

Digital literacy is the ability to access, manage, understand, integrate, communicate, evaluate and create information safely and appropriately through digital devices and network technologies for employment, decent jobs and entrepreneurship. It includes competences that are variously referred to as computer literacy, ICT literacy, information literacy and media literacy.

As with the other GAML task forces, the measurement strategy tackles, in turn, questions of relevance, implementation and interpretation *(see **Table 5.1**).* Two steps are being implemented in 2017/2018 and 2018/2019 with respect to relevance. The first step was to develop a content framework. While there was no globally-agreed framework, there were national or cross-national competence frameworks already developed:

- Notably the Digital Competence Framework for Citizens (DigComp) of the European Commission with 5 competence areas and 21 competences *(see Section 5.2);* and
- Specifically for assessments, the IEA ICILS and the OECD PIAAC, of which only the latter targets adults, which is the focus group of Target 4.4.

Given that DigComp is a comprehensive framework for youth and adults developed over several years and in consultation with several countries, it was an attractive point of departure. The key question was whether it was relevant not only for high-income countries but also for the rest of the world. The first activity of the task force was to invite the Hong Kong University Centre for Information Technology in Education (CITE) to investigate what adjustments would be needed to DigComp *(see Section 5.3).*

The CITE team first found information on digital literacy frameworks in 47 countries. It then mapped the competence areas of six national (Canada, Chile, Costa Rica, India, Kenya and Philippines) and

**Table 5.1 GAML Task Force 4.4 measurement strategy**

| Cross-national examples | | Global reporting<br>Standard expected GAML outputs | Task Force activities<br>2017/2018 | <br>2018/2019 | <br>2019/2020 |
|---|---|---|---|---|---|
| **Relevance** | | | | | |
| Assessment and competence frameworks | **IEA** International Computer and Information Literacy Study (ICILS)<br>**OECD** Programme for the International Assessment of Adult Competencies (PIAAC)<br>**ECDL Foundation** International Computer Driving License (ICDL)<br>**European Commission** Digital Competence Framework for Citizens (DigComp 2.1)<br>**LSE/Twente/Oii** Measuring digital skills | Has a learning assessment taken place?<br>■ Catalogue of learning assessments<br>What is the least common denominator?<br>■ Global content framework<br>How do different assessment frameworks map against the global content framework?<br>■ Content coding scheme<br>■ Evaluation of content alignment | <br>X<br><br>X | <br><br><br><br>X<br>X | <br><br><br><br> |
| **Implementation** | | | | | |
| Technical standards<br>■ Sample, coverage etc.<br>■ Modality, security etc. | | Are the assessments technically robust?<br>■ Evaluation of data quality | | | X |
| **Interpretation** | | | | | |
| ■ Reporting scale<br>■ Performance levels<br>■ Benchmarks | European Union Digital Economy and Society Index (DESI)<br>Dimension 2: Human capital/digital skills | How does learning improve?<br>■ Learning progression<br>A score that is attached to each learning level<br>■ Reporting scale<br>What level should learners achieve on that scale?<br>■ Minimum proficiency level | | | <br>X<br><br>X<br><br>X |

*Source: GAML Task Force 4.4.*

three popular enterprise (IC3, ICDL and Microsoft) frameworks onto DigComp 2.1 and found two types of competence areas that were qualitatively different from those defined in DigComp *(see **Table 5.2**)*:

- A competence area which would capture familiarity with basic operations of digital devices that are usually taken for given in rich countries; and
- A cross-cutting competence area, which would refer to specific careers or career opportunities.

The cross-cutting competence area is defined through everyday-use examples that drew on different cultural, economic and technological settings in low- and middle-income countries and four economic sectors: agriculture, energy, finance and transportation.

**Table 5.2 Competence areas and competences of the Digital Literacy Global Framework**

| Competence area | Competences |
|---|---|
| 0. Fundamentals of hardware and software | 0.1 Basic knowledge of hardware such as turning on/off and charging, locking devices<br>0.2 Basic knowledge of software such as user account and password management, login, and how to do privacy settings, etc. |
| 1. Information and data literacy | 1.1 Browsing, searching and filtering data, information and digital content<br>1.2 Evaluating data, information and digital content<br>1.3 Managing data, information and digital content |
| 2. Communication and collaboration | 2.1 Interacting through digital technologies<br>2.2 Sharing through digital technologies<br>2.3 Engaging in citizenship through digital technologies<br>2.4 Collaborating through digital technologies<br>2.5 Netiquette<br>2.6 Managing digital identity |
| 3. Digital content creation | 3.1 Developing digital content<br>3.2 Integrating and re-elaborating digital content<br>3.3 Copyright and licences<br>3.4 Programming |
| 4. Safety | 4.1 Protecting devices<br>4.2 Protecting personal data and privacy<br>4.3 Protecting health and well-being<br>4.4 Protecting the environment |
| 5. Problem solving | 5.1 Solving technical problems<br>5.2 Identifying needs and technological responses<br>5.3 Creatively using digital technologies<br>5.4 Identifying digital competence gaps<br>5.5 Computational thinking |
| 6. Career-related competences | 6. Career-related competences refers to the knowledge and skills required to operate specialized hardware/software for a particular field, such as engineering design software and hardware tools, or the use of learning management systems to deliver fully online or blended courses. |

**Note:** These competences draw on the DigComp 2.1 competences. Underscored competence areas and competences are additions to DigComp 2.1.
*Source: UIS, 2018c.*

They were sourced from news articles, videos, non-governmental organization (NGO) reports, software applications and company websites. For instance, three agricultural examples of increasing complexity referred to farmers making better farming and trading decisions using a mobile phone service, buying and selling products through a smart phone app, and building a data-driven irrigation system using moisture sensors linked to a laptop.

Ultimately, grounding digital literacy competences, proficiency levels and assessments in examples of use, and not at the conceptual level in frameworks, can show a contextualised approach to digital literacy

competence achievement and can result in multiple pathways to achieve digital literacy in a given country. The developmental context determines the pathway to digital literacy, and countries can make decisions to show progress towards digital literacy depending on this context. This approach helps address the issue of relevance across countries but does not resolve the issue of relevance over time in the face of rapid changes to technology and ICT uses.

### 5.1.2 Mapping the framework to existing assessments – and beyond

The second step in the GAML Task Force measurement strategy is to catalogue existing assessment tools and map them to the framework. The Centre for Educational Technology at Tallinn University is looking at different types of digital literacy assessments that vary by focus, application domain, purpose (e.g. admission, certification, training needs assessment and employment), target population, scale, item development, reliability and validity, mode of delivery, cost, scalability and accreditation *(see Section 5.4).*

The range of skills covered in digital literacy assessments is much wider than in assessments of reading and mathematics, which tend to follow a clearly-defined curriculum. In addition, digital literacy assessments vary in terms of the responsible authority. Non-government providers are more often involved in administering them. As a result, these assessments become proprietary and less transparent. Particular attention will be paid to the tool being launched by the European Commission.

The key expected result of the listing and mapping exercise will be recommendations regarding the types of existing assessments that hold the strongest potential for assessing the competences of the Global Digital Literacy Framework from the point of view of scope and methodology relative to the framework, technology requirements and delivery mode. Work could then begin in 2019/2020 on developing existing tools, where necessary, to accommodate the demands of global monitoring and introduce a reporting scale and proficiency level.

## 5.2 DIGCOMP: THE EUROPEAN DIGITAL COMPETENCE FRAMEWORK[64]

Being digitally competent is becoming a necessity for everyone to participate in our increasingly-digitalised economy and society. This is a major challenge for many countries, including those of the European Union (EU). According to the European digital skills indicator, 43% of the EU population and 35% of the EU labour force had an insufficient level of digital skills, and 17% and 10%, respectively, had no digital skills in 2017, mostly because they did not use the Internet. The construction of the composite indicator is based on DigComp, which was first published by the European Commission in 2013 as a reference framework to support the development of digital competence of individuals in Europe (European Commission, 2018a).

DigComp defines and describes which competences are needed today to use digital technologies in a confident, critical, collaborative and creative way to achieve goals related to work, learning, leisure, inclusion and participation in the digital society.

DigComp was developed by the Joint Research Centre of the European Commission as a scientific project, initially on behalf of the Directorate General for Education and Culture and, more recently, on behalf of the Directorate General for Employment, Social Affairs and Inclusion. In order to produce the framework, extensive literature review, case study research and stakeholder consultation processes were carried out. More than 200 experts and a variety of stakeholders from Europe have been involved in developing DigComp. Updates and further elaborations of the framework were carried out in June 2016 (DigComp 2.0) and May 2017 (DigComp 2.1).

64 Written by Yves Punie, Riina Vuorikari and Marcelino Cabrer, Joint-Research Centre, European Commission. The views expressed in this paper are those of the authors and should not be attributed to the European Commission.

Currently, there is a great variety of DigComp practices across Europe demonstrating the many opportunities it offers for different aims in digital competence initiatives, including digital skills goal-setting and strategy design, the development of education and training programmes, competence assessment and recognition. These initiatives take place in various domains, including formal education and training, lifelong learning and employment for a wide range of stakeholders addressing different target groups, such as students, workers and jobseekers.

Primary stakeholders of DigComp are education and training policymakers at regional and national levels, educational and training experts and organizations, research and support agencies, employers and recruiters, economic development professionals, public administrators, professional associations and private firms. Digital competence initiatives by students, citizens, workers, small entrepreneurs, teachers and educators may also benefit from this work. Stakeholders report that the value of using the DigComp framework relates to its character as a European framework, its contribution to establishing a common language and framework for understanding of digital competence, the quality and flexibility of the framework and its guiding function for education and training actions.

### 5.2.1 What is DigComp?

Being digitally competent is more than being able to use the latest device or software. Digital competence is a key, transversal competence, emphasising the ability to use digital technologies in a critical, collaborative and creative way. DigComp is a conceptual reference model intended to support a comprehensive understanding of digital competence in everyday life, particularly learning. DigComp presents five competence areas which outline the key components of digital competence.

i) **Information and data literacy:** required to articulate information needs, to locate and retrieve digital data, information and content, to judge the relevance of the source and its content, and to store, manage and organize digital data, information and content.

ii) **Communication and collaboration:** required to interact, communicate and collaborate through digital technologies and to manage one's digital identity and reputation, while being aware of cultural and generational diversity. Required to participate in society through digital services and participatory citizenship.

iii) **Digital content creation:** required to create and edit digital content and to improve and integrate information, while understanding how copyright and licences are to be applied. Required to know how to give understandable instructions for a computer system.

iv) **Safety:** required to protect devices, content, personal data and privacy, physical and psychological health and social well-being; required to be aware of the environmental impact of digital technologies and their use;

v) **Problem-solving:** required to identify needs and problems and to resolve conceptual problems and problem situations in digital environments, to use digital tools to innovate processes and products, and to keep up-to-date with the digital evolution.

In detail, DigComp sets out 21 competences that are described across eight proficiency levels through learning outcomes, from the most basic level to highly-specialised levels. Since DigComp has been designed to be a reference framework for digital competence, the framework is descriptive rather than prescriptive, highlighting the importance of all competences. Further elaboration of the content and the level of the competences can be done by users, which makes the framework flexible and adaptable. Some effort may be required to adapt DigComp content to local goals and specific circumstances. The question of digital skills must be embraced consistently across the sectors and actors involved in education, training, support, employment and development.

**Figure 5.2 How to swim in the digital ocean**

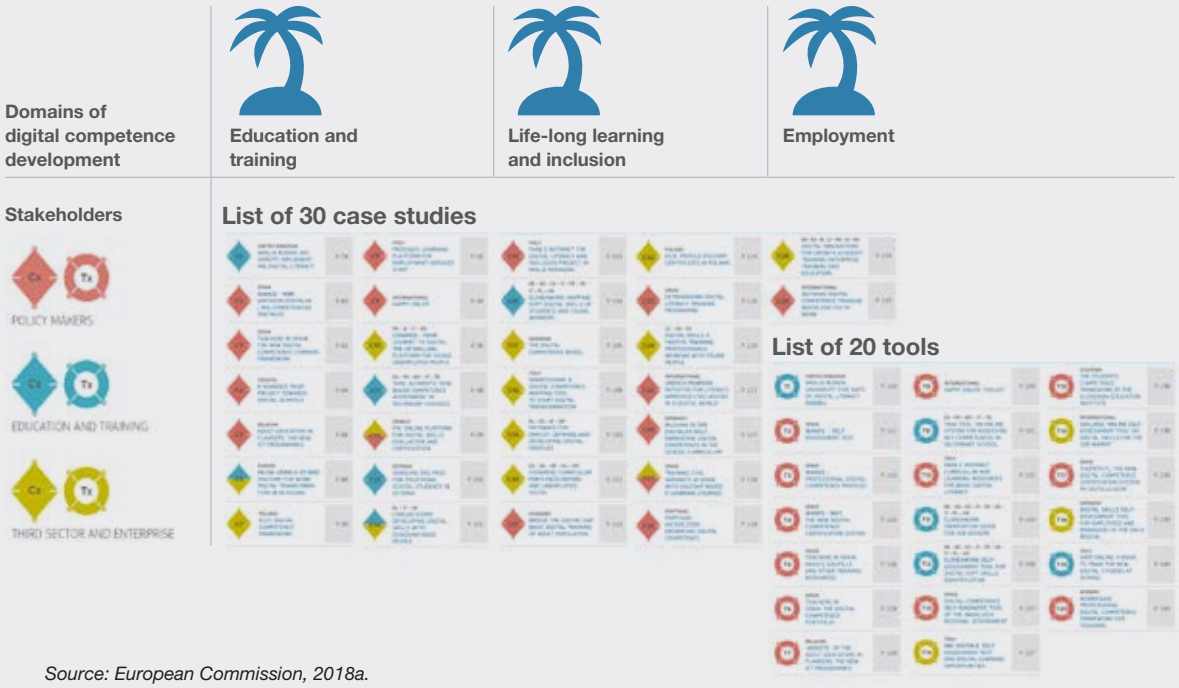

*Source: European Commission, 2018a.*

## 5.2.2 Uptake of DigComp

The European Commission has prioritised and supported the development of digital skills through a range of policies and actions, working with Member States in supporting learners, employees, jobseekers and innovators in every setting. Digital competence is confirmed to be one of the eight key competences for lifelong learning following the adoption of the Council Recommendation on Key Competences for Lifelong Learning in May 2018. DigComp represents a milestone of this journey; it is now regarded as the seminal contribution for the development of a Global Digital Literacy Framework, as proposed by UNESCO, within the context of the SDGs and GAML (UNESCO, 2018).

DigComp is widely used across EU Member States and beyond as a versatile tool to support digital competence building *(see **Figure 5.2***). The recently-published user guide, *DigComp into Action — Get Inspired, Make It Happen* (European Commission, 2018a), provides an overview of DigComp practices. The guide demonstrates the inspiring level of use of DigComp to date across diverse sectors, and it highlights an important message: digital skills are relevant to every aspect of our lives. DigComp is being used and adapted by stakeholders across Europe to enable people to acquire the digital skills they need for participation in the workplace and to play an active role as confident citizens. The guide offers inspiration for using DigComp by providing a comprehensive overview of 30 examples describing their aims, achievements and the benefits and challenges of using the reference framework. Several overviews are offered to find the examples that may be of interest to the reader. The guide also sets out steps for implementation and use. The open participatory process underlying DigComp's production and its public documentation is broadly appreciated by stakeholders.

**Figure 5.3 DigComp structure and components**




*Source: European Commission, 2018a.*

The guide also provides links to resources and tools developed by stakeholders; for instance, it can support translating and adapting the framework to local contexts, by either addressing digital skills needs of intermediaries (teachers, trainers, youth workers, employment services, e-facilitators) or targeting individuals directly (jobseekers, workers, entrepreneurs). Digital competence training materials and self-assessment instruments are also developed. Work is also done to describe and detail digital competence professional profiles for certain professions. These include museum, library and university staff, civil servants, virtual office workers and professionals in industry 4.0, a current trend of automation in manufacturing.

The DigComp user guide aims to support the implementation process of DigComp, offer an opportunity to learn from each other and share a pool of available resources in different languages so that interested stakeholders can avoid starting DigComp initiatives from scratch *(see **Figure 5.3**)*.

### 5.2.3 DigComp learning outcomes

DigComp maps out four broad proficiency levels: foundation, intermediate, advanced and highly-specialised. These four levels can be further elaborated on by breaking them into eight levels, offering a more detailed description of progression criteria *(see **Figure 5.4**)*. The eight levels provide the granularity needed to develop learning materials, assess and recognise learning progression and describe tasks and competences in detail. Each of the eight levels represents a further step by the citizens in three domains: the acquisition of the competence according to its cognitive challenge, the complexity of the tasks they can handle and their autonomy in completing the task. For each competence, eight proficiency levels are defined. Each one is written out as a learning outcome containing knowledge, skills and attitudes outlined in one single descriptor (8 proficiency levels × 21 competences = 168 learning outcomes).

**Figure 5.4 Main keywords describing DigComp proficiency levels**

T.4 Main keywords that feature the proficiency levels

| 4 OVERALL LEVELS | Foundation | | Intermediate | | Advanced | | Highly specialised | |
|---|---|---|---|---|---|---|---|---|
| 8 GRANULAR LEVELS | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| COMPLEXITY OF TASKS | Simple task | Simple task | Well-defined and routine tasks, and straightforward problems | Tasks, and well-defined and non-routine problems | Different tasks and problems | Most appropriate tasks | Resolve complex problems with limited solutions | Resolve complex problems with many interacting factors |
| AUTONOMY | With guidance | Autonomy and with guidance when needed | On my own | Independent and according to my needs | Guiding others | Able to adapt to others in a complex context | Integrate to contribute to the professional practice and to guide others | Propose new ideas and processes to the field |
| COGNITIVE DOMAIN | Remembering | Remembering | Understanding | Understanding | Applying | Evaluating | Creating | Creating |

*Source: European Commission, 2018a.*

The proficiency levels were inspired by the structure and vocabulary of the European Qualification Framework (EQF) and were defined as learning outcomes using action verbs following Bloom's taxonomy.

For instance, an individual at proficiency Level 2 is able to remember and carry out a simple task, with help from somebody only when they need it. However, a person at proficiency Level 5 can apply knowledge, carry out different tasks, solve problems and help others to do so as well.

### 5.2.4 Further work on digital competence frameworks

Further Joint Research Centre (JRC) work on DigComp, in collaboration with the European Commission's Directorate General for Employment, Social Affairs and Inclusion, will consist of maintaining its support to stakeholders and continuing to document and analyse its uptake and use. The development of a reliable and validated self-assessment instrument for DigComp will also be explored. Additionally, an analysis of the further applicability of DigComp to employability settings is planned. The latter is aimed to provide labour market intermediaries, such as public employment services, with concrete tools for tackling skills mismatches and, on the other hand, up-skilling and re-skilling opportunities for individuals and sectors most in need of digital skills.

Digital competence development is also crucial for educators and educational organizations. The JRC published the Digital Competence Framework for Educators (DigCompEdu) at the end of 2017, and its purpose is to describe and define what it means for educators at all levels to be digitally competent. It provides a general reference framework to support educator-specific digital competences in Europe. It consists of 22 competences for teaching in a digital society along six competence areas. This work is now continued with the development of an assessment instrument for DigCompEdu.

DigCompOrg is a comprehensive and generic conceptual framework that reflects all aspects of the process of systematically-integrating digital learning in educational organizations from all education sectors. The conceptual model was published by the JRC in 2015. It contains 7 key areas and 74 specific descriptors on digital age learning. While DigCompOrg is for all educational organizations, a

specific tool for schools, the self-reflection tool for digitally-capable school (SELFIE), became available in 2018. SELFIE is a free, online application that schools in Europe and beyond can use to self-reflect on their level of digital capacity in order to develop an improvement plan. Both DigCompEdu and DigCompOrg's SELFIE are developed by the JRC in collaboration with the European Commission's Directorate General for Education and Culture.

Together, the three frameworks provide a comprehensive approach to capacity building for the digital transformation of education and training in Europe and the world at large. DigCompOrg is geared at educational organizations, DigCompEdu targets educators' digital capacity, and DigComp addresses citizens, students, workers and intermediaries such as employment agencies. DigComp can help in bridging the digital divide, thus contributing to the 2030 Agenda for Sustainable Development. It can be adapted to measure and increase the proficiency level of citizens' digital skills and therefore foster greater proliferation of digital literacy, an explicit aim of SDG 4.

## 5.3 A GLOBAL FRAMEWORK OF REFERENCE ON DIGITAL LITERACY SKILLS FOR SDG INDICATOR 4.4.2 [65, 66]

CITE of the University of Hong Kong was commissioned by the UIS in November 2017 to conduct a study to develop a Digital Literacy Global Framework (DLGF) to serve as a foundation for the further development of thematic indicators under the SDGs. The DLGF is specifically related to Indicator 4.4.2, which tracks the "percentage of youth/adults who have achieved at least a minimum level of proficiency in digital literacy skills", and is one of the three indicators for SDG Target 4.4 that aims to substantially increase the number of youth and adults who have relevant skills, including technical and vocational skills, for employment, decent jobs and entrepreneurship.

### 5.3.1 Project methodology

The project was conducted in three phases: i) a synthesis of existing regional, national and sub-national frameworks to identify skills and competences relevant for the global context and, in particular, analysing the extent to which existing, well-developed and all-encompassing frameworks would be relevant for all countries, whether rich or poor and over time; ii) an in-depth consultation with education experts from different regions; and iii) an online consultation through a survey involving experts from Member States and UN entities.

In conducting this study, we have used DigComp (Vuorikari et.al., 2016) as the initial framework. This was developed on the basis of comprehensive reviews of literature and policy documents, as well as extensive consultations with different stakeholders in Europe *(see Section 5.2).* However, as the value of having a DLGF lies in its meaningfulness to different socioeconomic and developmental contexts, we have made particular efforts to include materials and experts from countries outside Europe and North America, where digital literacy policies and provisions are less developed. Those countries that are most likely to benefit from a DLGF will also not have well-developed policies or research literature related to digital literacy. We have thus added to our Phase 1 work the identification and analysis of digital literacy competences as illustrated by the use of ICT in major socioeconomic sectors, particularly in developing countries.

### 5.3.2 Project findings

There are many national and regional efforts to develop and implement digital literacy frameworks and strategic plans to bolster citizens' digital literacy. However, there are differences regarding the definitions for digital literacy and the purposes such frameworks were intended to serve. Some

65 Written by Nancy Law, University of Hong Kong. David Woo contributed greatly to the development of this framework as the project manager for the study. Other project team members include Jimmy de la Torre and Gary Wong at the Faculty of Education, University of Hong Kong.

66 The full report on this study can be downloaded from http://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf. Interested readers can also find more information about the study methodology, instruments and exemplars for the Digital Literacy Pathways Mapping Methodology from the project website: http://gaml.cite.hku.hk/

consider digital literacy as a new literacy comprising multiple dimensions and represented in new, multimodal social practices, and is greater than the sum of the other literacies (Ala-Mutka, 2011). The proposed DLGF is intended to serve as the basis for monitoring, assessing and further developing digital literacy across a wide variety of socioeconomic and developmental settings. Hence, the resulting framework needs to be capable of being operationalised to serve this purpose.

In reviewing related frameworks collected from government and non-government agencies, frequently used terms included access, manage, understand, integrate, communicate, evaluate and create. Using these ideas as a foundation, we propose the following definition for digital literacy:

Digital literacy is the ability to access, manage, understand, integrate, communicate, evaluate, and create information safely and appropriately through digital technologies for employment, decent jobs, and entrepreneurship. It includes competences that are variously referred to as computer literacy, ICT literacy, information literacy, and media literacy.

In the remainder of this section, we report the project findings, following the order in which the research tasks were conducted.

### Mapping of existing regional, national and sub-national frameworks

We conducted a systematic search for digital literacy frameworks in the targeted regions and countries using country names in combination with search terms. These terms included digital, literacy, competences, skills, ICT, computer and information. The goal was not to have a statistically-representative collection of existing frameworks but to identify as broad a range of features in the frameworks as possible. As DigComp 2.0 already reflects the full range of digital literacy competences that are found to be important in Europe and other developed western countries, we have focused our search on countries in other regions.

Our search found information about specific digital literacy frameworks in 47 countries, with the following regional distribution: Asia (11), EU (2), high-income countries outside the EU (2), Latin America (5), Middle East and North Africa (12), sub-Saharan Africa (13) and other regions (2). A key limitation to the search results is that these are limited to information accessible through the English language.

Our analysis found that some countries have multiple frameworks in use, often for different purposes. In some countries, particularly in economically less-developed ones, enterprise digital literacy frameworks[67] developed by commercial entities that offer training courses and certification have been adopted for the purpose of human resource development and qualification requirements for jobs. While these frameworks do not have official status as a national framework, they play an important role in influencing digital literacy development in the respective contexts.

Of the 47 countries with frameworks, 11 have developed their own national frameworks; of these, 7 have adopted enterprise frameworks. At the same time, 36 of these countries only have enterprise frameworks and some have adopted more than one enterprise framework. Therefore, multinational commercial enterprises have a major role in influencing the digital literacy competences that are being taught and assessed, particularly in developing countries.

In mapping the competences in the collected frameworks to the DigComp 2.0 framework, we find two competence areas that are not explicitly included in the latter (the full list of competence areas are listed in **Table 5.3**):

- *Devices and software operations (CA0)* – basic operations of digital devices, understanding basic

67 We have identified three digital literacy enterprise frameworks adopted by the 47 countries in our study, in decreasing order of popularity (note that some countries adopt more than one framework):
International Computer Driver's Licence (ICDL)—adopted in 31 countries;
Certiport Internet and Computing Core Certification (IC³)—adopted in 13 countries; and
Microsoft Digital Literacy Standard Curriculum—adopted in 11 countries.

**Table 5.3 Competence areas and competences for the proposed DLGF**

| Competence area (CA) | Competences |
|---|---|
| CA0. Devices and software operations | 0.1 Physical operations of digital devices<br>0.2 Software operations in digital devices |
| CA1. Information and data literacy | 1.1 Browsing, searching and filtering data, information and digital content<br>1.2 Evaluating data, information and digital content<br>1.3 Managing data, information and digital content |
| CA2. Communication and collaboration | 2.1 Interacting through digital technologies<br>2.2 Sharing through digital technologies<br>2.3 Engaging in citizenship through digital technologies<br>2.4 Collaborating through digital technologies<br>2.5 Netiquette<br>2.6 Managing digital identity |
| CA3. Digital content creation | 3.1 Developing digital content<br>3.2 Integrating and re-elaborating digital content<br>3.3 Copyright and licences<br>3.4 Programming |
| CA4. Safety | 4.1 Protecting devices<br>4.2 Protecting personal data and privacy<br>4.3 Protecting health and well-being<br>4.4 Protecting the environment |
| CA5. Problem solving | 5.1 Solving technical problems<br>5.2 Identifying needs and technological responses<br>5.3 Creatively using digital technologies<br>5.4 Identifying digital competence gaps<br>5.5 Computational thinking |
| CA6. Career-related competences | 6.1 Operating specialised digital technologies for a particular field<br>6.2 Interpreting data, information and digital content for a particular field |

*Source: UNESCO Institute for Statistics (UIS), 2018c.*

concepts of hardware and software, and operating a graphical user interface.

- *Career-related competences (CA6)* – use of digital technologies that are important productivity tools for particular business sectors, such as learning management systems (education and training), computer-aided design (architecture and engineering), and social media (marketing). This competence area is included in two of the three enterprise frameworks we identified.

A comparison of the frequency of coverage across competences shows the most frequently-included competences to be *Devices and software operations* and *Information and data literacy*, and the least popular was *Protecting the environment.*

### Mapping of digital literacy competences in examples of digital technology use

To provide meaningful guidelines for the provision of training, monitoring and assessment of digital

literacy associated with employment, decent jobs and entrepreneurship in diverse contextual settings, we searched news and media reports to identify examples of digital literacy use in: i) everyday contexts in four key economic sectors (agriculture, energy, finance and transportation) in a wide range of countries outside Europe; and ii) the empowerment of communities suffering from systemic economic, social and political vulnerabilities, such as high levels of low-skilled and illiterate women in poor communities and displaced populations such as refugees. There were two steps in the mapping process. First, we identified the functional operations that the user may need to perform in each of the tasks in an example. This resulted in a total of 15 across the collected examples. We then mapped each function to the competence framework resulting from the framework mapping exercise described above.

There are several key findings from this process:

- The 15 functions fall into two categories: general operations and financial transactions.
- All the examples do not require the use of a computer but require a network-enabled device, such as a mobile phone or a smartphone/tablet connected to the Internet.
- The digital literacy competence levels required for achieving the same function are dependent on the type of device used. For example, searching for goods and services and comparing prices differ greatly depending on whether a mobile phone or a smartphone is used.
- The digital literacy competences and the proficiency levels required on a smartphone are higher than on a mobile phone and quite different from a stand-alone computer. This also implies that the digital literacy competences and the proficiency levels achieved through training are dependent on the nature of the devices used.
- Not all the competences found in the framework mapping exercise were found in the analysis of the examples used, indicating that digital literacy competences required in everyday uses would be narrower than those required in specialised situations such as in employment.
- The specific digital literacycompetences and proficiency levels that are important, as well as the opportunities to learn such competences, depend on their specific country and economic sector contexts, including the technology and Internet infrastructure and access available in the community.

Based on the finding and mapping methodology developed, we have developed a *pathway mapping methodology* to guide countries, sectors, groups and individuals to develop strategies and plans for advancing their own digital literacy development goals and pathways. A pathway here refers to the *digital development pathway* that individuals, groups, communities or sectors intend to pursue in terms of digital technology adoption/integration to order to achieve the developmental goal(s) targeted. For example, a farmer's digital development pathway could be to move from using a mobile phone to seek better offers for produce, to using a smartphone to seek better market intelligence as well as direct channels of reaching customers. A digital literacy development pathway for this farmer comprises the differences in digital literacy competences required for the use case scenario he/she aspires to (the smartphone use scenario described) with the set of digital literacy competences he/she possesses for the current farming-related activities. In general, a digital literacy development pathway can be constructed through identifying the differences in digital literacy competences between the current digital technology-use scenario with a scenario targeted for developmental purposes.

### In-depth consultation

As part of the in-depth consultation phase, experts were invited to review the draft executive summary of a DLGF. This was followed by an online interview to seek their feedback on the relevance of digital literacy in their local contexts and the suitability of the proposed DLGF. The consultation was completed

by 15 experts, representing coverage of at least two countries from each of the six targeted regions: Africa, Asia, the EU, high-income nations outside the EU, Latin America, and the Middle East and North Africa. Some experts have served in projects covering multiple countries and regions. The findings show:

- There is general agreement about the relevance of the proposed framework, even though there are some opposing views, primarily from experts from the economically-developed countries, about the competence areas *Devices and software operations* and *Career-related competences* that have been added to the DigComp 2.0 framework.
- When consulted on whether there is any missing digital literacy competences in the framework, computational thinking came up most frequently. The view from these experts is that computational thinking is the application of algorithmic thinking as an integral part of problem-solving competences in a digital world. This may not involve programming in specific computer languages; therefore, this is different from programming as a method of digital content creation.
- The proposed pathway mapping methodology is found to be helpful for developing digital literacy strategies and plans suited to specific contexts and needs. Some experts provided further examples of digital literacy application that can be used to develop such pathways, but also foresaw difficulties in implementing the pathway mapping methodology.

#### Online consultation

For the online consultation, respondents were asked to review a short video presentation on the proposed DLGF before completing a 22-item survey on the competence areas and competences in the proposed DLGF, the pathway mapping methodology and background information about the respondent. To solicit input from a larger number of stakeholders from different countries, the online consultation was promoted through social media and research information management systems. A total of 31 complete responses was received at the end of the consultation period. The findings were very similar to those from the in-depth consultation.

### 5.3.3 Digital Literacy Global Framework proposed for Indicator 4.4.2

Based on the findings from both the in-depth and online consultations, the project team proposed a final version of the DLGF to the UIS for consideration, presented in **Table 5.3**.

It is important to note that different levels of proficiency can be associated with each competence. Both the competence area and the associated minimum proficiency level required for competent performance are dependent on the contexts of use involved. A digital literacy framework thus provides a basis for the further development of descriptors for different levels of proficiency for each of the competences. DigComp 2.1 (Carretero, Vuorikari and Punie, 2017) provides a good example of how descriptors for different levels of proficiency can be further developed based on a comprehensive digital literacy framework.

### 5.3.4 Recommendations for the next steps

The results from the research and consultation processes show that there is wide recognition of the value of a global framework to guide the development of digital literacy. Experts and stakeholders across diverse economic and regional contexts have generally agreed on the proposed DLGF and pathway mapping methodology, but the priorities for digital literacy development will differ depending on the context. Our findings also show that the DigComp 2.0 framework is a valuable and suitable basis for the development of a DLGF. The proposed framework and pathway mapping methodology can serve as a foundation for the development of: i) specific thematic indicators for Indicator 4.4.2; and ii) digital literacy frameworks, curricula and assessments in different countries and regions.

We further recommend that the proposed DLGF could serve as a lever for scaffolding inter-organizational coordination and collaboration on the enhancement of digital literacy development. In particular, collaboration on the implementation of the pathway mapping methodology to generate digital literacy training and assessment programmes may provide a fertile context for collaboration among entities in diverse socio-political and economic contexts using the DLGF as a common framework.

## 5.4 TOWARDS A NEW FRAMEWORK AND TOOL FOR ASSESSING DIGITAL LITERACY SKILLS OF YOUTH AND ADULTS (INDICATOR 4.4.2)[68]

This section discusses the approach and methodological challenges in an ongoing desk research project that aims to advise the UIS in designing an instrument to assess digital literacy skills in the context of collecting data on Indicator 4.4.2. The SDG Target 4.4 contains three indicators (UIS, 2018c):

- 4.4.1 Proportion of youth and adults with information and communications technology (ICT) skills, by type of skill.
- 4.4.2 Percentage of youth/adults who have achieved at least a minimum level of proficiency in digital literacy skills.
- 4.4.3 Youth/adult educational attainment rates by age group, economic activity status, levels of education and programme orientation.

The UIS is responsible for the development and validation of new methodologies for indicators under SDG Target 4.4. While Indicators 4.4.1 and 4.4.3 have already been implemented in reporting for 2017, the status of Indicator 4.4.2 is still under development (UIS, 2018c). Although many countries have been collecting data on digital skills or ICT literacy of their citizens for various purposes, there is no common agreement on what constitutes a minimum or basic level of proficiency in digital literacy that would allow aggregation of national data on the global level. As a result, there is a serious knowledge gap about the global state of digital literacy skills of youth and adults, while these skills play an increasingly important role in achieving SDG 4.

There have been some supra-national initiatives in this field, but those have focused on international assessments within a few countries (e.g. ICILS or ICDL). All these supra-national initiatives could definitely inform the UIS in designing a global instrument for collecting reliable and valid data on the digital literacy target, but none of these practices was specifically designed to inform Indicator 4.4.2.

The UIS should also keep an eye on the development of supra-national policy indicators on digital literacy. The EC has defined a new standard on a digital competence framework for citizens (DigComp, see Section 5.2), which has already been used for various purposes in several European countries (Carretero et.al., 2017). DG Connect and Eurostat have already used DigComp to redesign their digital skills indicator in 2015. Their survey asks respondents about digital activities carried out within the previous three months, assuming that "persons having realised certain activities have the corresponding skills" (European Commission, 2016). The indicator defines three levels of proficiency: below basic, basic and above basic levels. However, there is no common European instrument for performance-based assessment of digital competence of citizens based on DigComp.

As a major milestone in the process of developing its framework for digital literacy, the UIS commissioned a report, *A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2* (UIS, 2018c). This report reviews digital literacy assessment frameworks used in 47 countries and summarises consultations with a number of experts, resulting in the suggestion to use the European DigComp framework as the foundation for the UIS DLGF, while expanding it by five additional competences and adding two competence areas. The report raises three challenges. First, the

68 Written by Mart Laanpere, Senior Researcher, Centre for Educational Technology, Tallinn University.

need for mapping existing instruments for digital skills assessment to DLGF, pointing out that "...there is not a one-size-fits-all assessment of digital competence that can serve all purposes and contexts". Second, it also calls for cost-effective cross-national R&D programmes to develop and validate "context-sensitive and fit-for-purpose digital literacy indicators and assessment instruments". Third, the report points out the discrepancy between the proficiency levels and related measurement scales of the SDG indicator versus DigComp. While Indicator 4.4.2 focuses on a minimum level of proficiency, DigComp distinguishes eight proficiency levels.

These three challenges raised by the authors of the report are addressed by ongoing desk research that has three objectives:

- Mapping existing digital literacy assessments to DLGF;
- Evaluating advantages and disadvantages of selected assessments that cover a large part of the DLGF, with emphasis on their cost-effectiveness for rollout on a population scale; and
- Recommending the next steps on developing an assessment tool suitable for Indicator 4.4.2.

### 5.4.1 Methodological challenges in the assessment of digital literacy

Digital literacy is a relatively new concept to join competing concepts such as ICT, media, information and computer literacy (or competence). Ferrari (2013) was among the first authors who tried to settle the relationship between these existing labels and newcomers (digital literacy/competence) in a similar manner with the definition suggested by authors of the 2018 UIS report: "Digital literacy is the ability to access, manage, understand, integrate, communicate, evaluate and create information safely and appropriately through digital technologies for employment, decent jobs and entrepreneurship. It includes competences that are variously referred to as computer literacy, ICT literacy, information literacy and media literacy".

This definition builds on previous practices by incorporating vocabulary from predecessors (e.g. from information, media and ICT literacy frameworks), resulting in a list of 26 competences grouped into seven competence areas. As experience with the EC's DigComp has demonstrated, such a competence framework can be used for various pragmatic purposes: re-designing the outdated curricula and professional development programmes, developing policy indicators, professional accreditation, recruitment and (to a lesser extent) research.

As an alternative to this pragmatic approach, recent psychometric approaches to measuring digital literacy have been guided by Multidimensional Item Response Theory (MIRT) that understands Computer and Information Literacy (Fraillon et al., 2014) or Digital Information Literacy (Sparks et.al., 2016) as a single latent trait that cannot be directly observed in test situations and, thus, should be inferred indirectly through statistical analysis of test results. Like any mathematical model, MIRT has some assumptions that need to be fulfilled in order to make valid inferences on the basis of test results. For instance, the monotonicity assumption requires that the instrument does not make knowledgeable persons more likely to participate in the test (Chenery and Srinivasan, eds., 1988). An assumption of local independence means that performance in one item in a test does not influence performance in other items. While such assumptions are relatively easier to guarantee in the case of knowledge-based multiple choice tests, the same might be quite difficult in the case of authentic performance-based assessments.

Two approaches to digital literacy assessments that were described above illustrate the struggle between internal and external validity in the context of educational assessments. Validity in general is understood as the degree to which test results can be interpreted and used according to the stated purposes of the assessment (AERA, 2014). Internal validity refers to methodological correctness/ coherence of a research instrument, while external validity can be interpreted as its re-usability through

the relevance or usefulness for a wider audience. The pragmatic approach to defining and measuring digital literacy tends to result in poorer internal validity but higher external validity of the assessment instrument, as it is better understood and accepted by various stakeholders (most of whom may not have a background in mathematical statistics or psychometry). On the other hand, the psychometric approach guarantees higher internal validity quite often at the expense of reduced external validity.

The UIS report (2018) recommends using pathway mapping methodology for operationalising a DLGF, focusing rather on users' perception of digital literacy in various contexts and concerning external validity of assessment. Eventually, the digital literacy assessment based on DLGF will have to address the challenge of balancing internal and external validity, both through methodological considerations and the design of the digital literacy assessment instrument.

### 5.4.2 Existing instruments for assessing digital literacy

Carretero et al. (2017) have reviewed 22 existing instruments that are used to assess digital competence in line with the DigComp framework in various European countries. They grouped these instruments into three major categories based on the data collection approach:

- **Performance assessment**, where individuals are monitored by a human observer or software while being engaged in solving authentic, real-life problems by using common software tools (e.g. browser, word processor, spreadsheet).
- **Knowledge-based assessment**, where individuals are responding to carefully designed test items that measure both declarative and procedural knowledge.
- **Self-assessment**, where individuals are asked to evaluate their knowledge and skills by means of questionnaires that might range from structured scales to free-form reflection.

These approaches can be strengthened by secondary data-gathering and analysis (e.g. by providing an e-portfolio that contains creative works, certificates and other documentary evidence). It is likely that performance assessment and analysis of secondary data are not cost-effective approaches in the context of global assessment of digital literacy in the context of the SDGs. Self-assessment would be the easiest and most cost-effective to implement but will likely suffer from low reliability and validity. However, it should be possible to combine self-assessment with knowledge-based or performance assessment. For instance, Põldoja et.al. (2014) have designed and validated an instrument called DigiMina that combined self-assessment of teachers' digital competence with peer-assessment, knowledge-based tests and an e-portfolio containing teacher's reflections and creative work. Within the DigCompEdu project, JRC tried to balance internal and external validity in assessing a school's digital capability with the design of the SELFIE tool, so that schools are allowed to expand the scientifically-validated core instrument with additional items from a pre-designed, publicly available pool or even design their own additional items that seem relevant to them (Joint Research Centre, European Commission,2018). The future instrument that will be designed by the UIS for digital literacy assessment might also benefit from a similar balancing of needs for global standardisation (contributing to internal validity) and local context (contributing to external validity).

The ongoing study uses the three categories of instruments for digital literacy assessment described by Carretero et al. (2017) to identify the existing practices and evaluate their applicability in the context of data collection for Indicator 4.4.2. The applicability analysis mainly focuses on the cost-effectiveness of the given instrument but also considers its reliability and validity, following the discussion above.

The existing digital literacy assessment practices and instruments will be sought from three types of sources:

- Scientific research publications;
- Policy documents in education and employment domains; and
- Professional certification frameworks and related technical documents.

The current study will map the existing assessments to DLGF and address the methodological challenges described in this chapter, resulting in recommendations to the UIS regarding the next steps in developing a new instrument for assessing Indicator 4.4.2 that is cost-effective, reliable and valid (both internally and externally).

# 6. Learning evidence and approaches to measure SDG functional literacy and numeracy

SDG 4 calls for an increased focus on learning outcomes, with five of the ten education targets highlighting learning skills and outcomes for children and adults. The UIS established GAML in 2016 as a platform to convene technical experts, donors and international organizations to provide technical solutions to the learning-related indicators. The GAML work programme includes the development of standards, guidelines and measurement tools to collect data to inform SDG 4 indicators.

Target 4.6 calls on countries to "ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy" by 2030. More specifically, Indicator 4.6.1 refers to: "Proportion of population in a given age group achieving at least a fixed level of proficiency in functional (a) literacy and (b) numeracy skills, by sex".

This chapter provides an overview of Indicator 4.6.1 and the strategy to improve reporting of the data. After presenting a framework to measure Indicator 4.6.1, the first section presents the various reporting options. Section 6.2 describes experience with the PIAAC, while Section 6.3 focuses on the World Bank initiative, STEP. The chapter ends with an analysis of RAAMA *(Recherche-action sur la mesure des apprentissages des bénéficiaires des programmes d'alphabétisation)*, which was initiated by the UNESCO Institute for Lifelong Learning (UIL).

## 6.1 FRAMEWORK FOR REPORTING INDICATOR 4.6

The fitness for use of any data system can only be evaluated against the overall purpose of the data. As documented in **Table 6.1**, comparative data on the level and distribution of adult literacy and numeracy skills are needed to serve five distinct purposes, which have implications for the data collection strategy.

Comparative data on literacy and numeracy are needed by multilateral and bilateral donors to guide their policies and programmes and to monitor progress towards international and national targets, including SDG Target 4.6. It is also imperative for countries to use the data to better understand their national situation.

Measures of literacy and numeracy need to be compared over time to determine relative needs and to track progress.

### 6.1.1 How Indicator 4.6 is informed to date

Currently, there are only two internationally-administered assessments, OECD's PIACC and the World Bank's STEP, which makes use of a version of PIAAC's literacy assessment. The UIS Literacy Assessment and Monitoring Programme (LAMP) has a methodological framework and tools that are relevant to low- and middle-income countries, though it is not currently being administered. This

**Table 6.1 Uses for data on literacy**

| Application type | General purpose | Related policy questions | Implication for data collection strategy |
|---|---|---|---|
| **Knowledge generation** | Identification of the causal mechanisms that link skills to outcomes | How do individuals acquire skills? How do they lose skills? How are skills linked to outcomes? | Needs longitudinal or repeated cross-sectional data with comparable measures of skills |
| **Policy and programme planning** | Planning government response to identified needs to meet social and economic goals | Which groups need to upgrade skills? How many people are in need? Where is need concentrated? | Needs profile of skills for key sub-groups |
| | Determination of funding levels | How much budget is needed to raise skills at the rate needed to achieve social and economic goals? | Need numbers of adults with different learning needs |
| **Monitoring** | Adjustment of policies, programmes and funding levels | Are skill levels rising at the expected rate? | Need repeated cross-sectional skills measures |
| | | Are skills-based inequalities in outcomes shrinking? | Need repeated cross-sectional skill measures for key sub-groups |
| **Evaluation** | Formal process to determine if programmes are performing as expected and meeting their objectives | Are government programmes effective and efficient? | Need data on skills gain/loss and costs for programme participants |
| **Administration** | Making decisions about specific units: individuals, regions, programmes | What criteria are applied to determine programme eligibility? | Need results that are reliable enough to keep Type I and Type II classification errors to acceptable levels |

*Source: UNESCO Institute for Statistics (UIS).*

assessment will be very valuable, since the tools and methodologies used to assess literacy and numeracy in high-income countries, like PIAAC, are considered inappropriate for lower-income countries.[69]

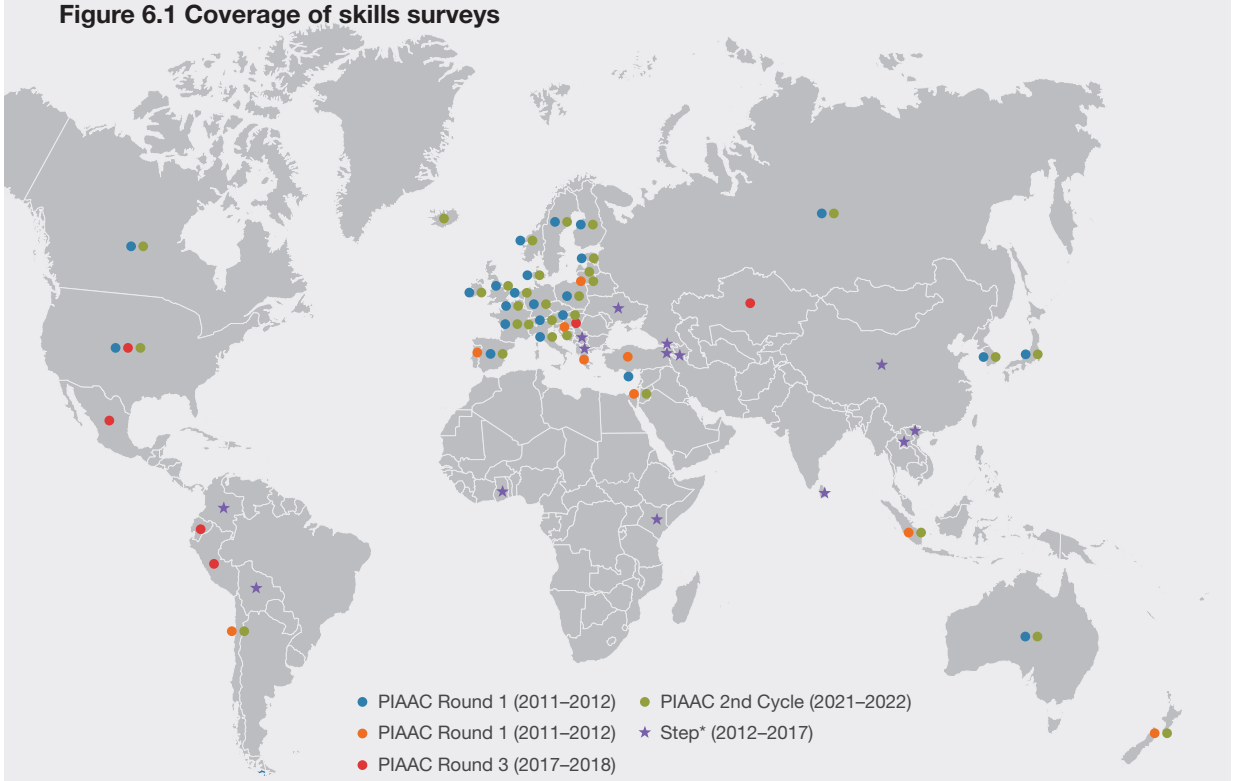
Conventionally, these assessments include:

- Administration of an extensive background questionnaire that identifies key population sub-groups, documents the determinants of skills differences and allows exploration of the impact that skill differences have on individual outcomes;
- Administration of a direct test of adult literacy and numeracy that covers the full range of skills in the population; and
- Administration of a direct test of the reading skills that support the emergence of fluid and automatic reading that characterise performance at the lower levels.

There are challenges and constraints associated with each of the two assessments. PIAAC tools may be relevant to the OECD or high-income countries,

69 An international report by the UIS (2017i) explores the differences between LAMP and the earlier versions of OECD's literacy assessments, International Adult Literacy Survey (IALS) and Adult Literacy and Life Skills Survey (ALL), and found that OECD literacy assessments that were conducted in OECD countries and exclusively in European languages do not address the challenges of testing in other contexts. LAMP was implemented between 2007 and 2008 in five countries as a pilot.

**Figure 6.1 Coverage of skills surveys**




**Note:** Population in urban centres.
*Source: UNESCO Institute for Statistics (UIS).*

but they are not relevant and might not be valid for low- and lower-middle-income countries. STEP tools were developed to target low- and middle-income countries. However, STEP focuses on work-relevant skills and does not measure numeracy. Its premise is that numeracy ability is highly correlated to literacy ability. However, using proxies as outcome variables can have a deleterious impact on measurement and behaviour (Gal, 2018). In low- and middle-income countries especially, it is possible to have respondents who are illiterate but have numeracy skills, so that correlation cannot be taken for granted. These assessment programmes are technically rigorous and respected, with many countries participating.

Of the international studies, the RAAMA study by UIL stands apart from the International Adult Literacy Survey (IALS), ALL, PIAAC, STEP and LAMP as the items used to assess literacy skills were not selected in a way to provide systematic coverage of the characteristics that underlie the relative difficulty of tasks, nor were results summarised using methods that confirm the stability, reliability and comparability of measurement. The RAAMA approach to measurement does not provide the needed cross-national comparisons of skills over time. However, the content framework developed for a low-literate population in literacy programmes may contribute to the development of the conceptual framework.

To date, only a handful of countries conduct national adult literacy assessments. Even though many have used the UNESCO definition of literacy as the basis for building their national adult literacy assessment, these assessments vary considerably in terms of content domain definitions and coverage. A number of national assessments were reviewed that measured literacy and numeracy skills indirectly. These assessments rely on self-reports of skills or on performance on very limited numbers of test items.

Research shows that these measures are unreliable, i.e. they are unable to support comparisons within or between countries (Niece and Murray, 1997). The fundamental problem with self-reports is that adult perceptions of their skill levels are conditioned by their use of their skills rather than their actual skill level and at times the social perception of having the skills. To make matters worse, the relationship of self-perceived skills to actual skills varies significantly among sub-populations within countries and across countries and over time. This renders these assessments of limited use to policymakers.

### 6.1.2 What are the challenges to report?

For SDG 4 monitoring and reporting, there is a need for a common definition and a common reference in reporting. In developing a strategy to monitor progress towards Target 4.6, the primary conceptual issue is agreement on the definitions and dimensions of the constructs of (adult) literacy and numeracy to be measured by Indicator 4.6.1. There are several main issues.

The indicator for Target 4.6 implies a need for measures:

i) of literacy and numeracy;
ii) that are statistically-representative of the adult population;
iii) that capture a range of definitions of functionality across countries;
iv) that can be compared under some criteria; and
v) that provide a set of cost-efficient options for countries.

The indicator specification also includes several subjective elements that require definition, including:

i) the definition of "functional" relative to literacy or numeracy;
ii) a menu of options for countries to measure and report; and
iii) a linking strategy to compare different options.

#### Definition of literacy and numeracy

The definition of literacy from the UN's Principles and Recommendations for Population and Housing Censuses, Revision 3 states:

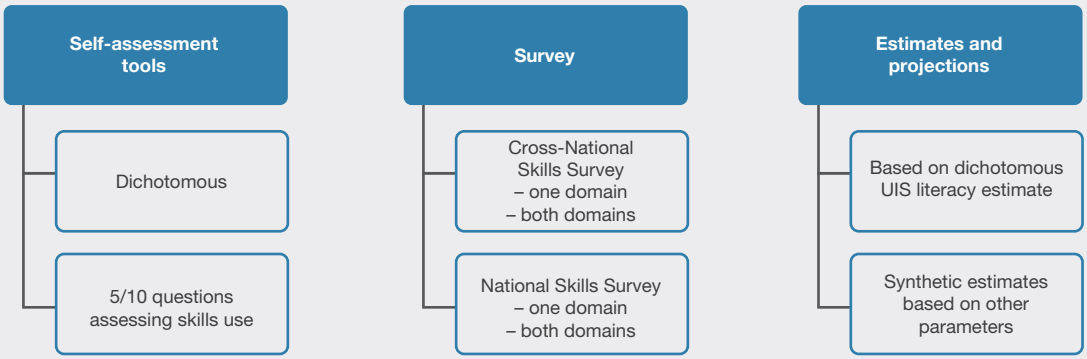
> *Literacy has historically been defined as the ability both to read and to write, distinguishing between "literate" and "illiterate" people. A literate person is one who can both read and write, with understanding, a short, simple statement on his or her everyday life. An illiterate person is one who cannot, with understanding, both read and write such a statement. Hence, a person capable of reading and writing only figures and his or her own name should be considered illiterate, as should a person who can read but not write as well as one who can read and write only a ritual phrase that has been memorized. However, a more modern understanding referring to literacy as a continuum of skills, levels, domains of application and functionality is now widely accepted. (UN, 2015).*

In the current generations of comparative assessments, functionality is defined as the level of literacy needed for an individual to cope with the demands that they confront in their daily lives and will differ by country and situation. For this reason, assessment has focused on the use of skills. Therefore each country must establish its own definition of what level constitutes the functional level(s) that reflects its definition of literacy skill-based inequality in individual outcomes, its targets for the performance of key social institutions, including firms and educational institutions, and its social and macro-economic demands.

#### No equivalent definition of numeracy exists

In terms of the conceptualisation of literacy and numeracy as a continuum, the situation in the field of adult assessments differs considerably from that of assessments of school-age children. The framework of the PIAAC assessment draws on a theoretical tradition that has underpinned the conceptualisation

**Figure 6.2 Summary of reporting options**




*Source: UNESCO Institute for Statistics (UIS).*

of literacy and, subsequently, numeracy in IALS, ALL and LAMP. Therefore, these assessments have a common conceptual framework.

The UNESCO definition of literacy was adopted by GAML with the conceptualisation of literacy and numeracy utilised by PIAAC with adaptations to extend the framework to include foundational skills. This was noted, even though the PIAAC conceptual framework is relatively comprehensive.

### 6.1.3 Exploring reporting options for Target 4.6.1

Given that the existing assessment tools (PIAAC and STEP) and data collection might be lengthy and costly, there are some alternatives that could be considered to report for Indicator 4.6.1 that include three broad categories: observed data based on self-assessment, administration of skills surveys and synthetic estimates *(see **Figure 6.2**)*.

#### Indirect and simplified measures

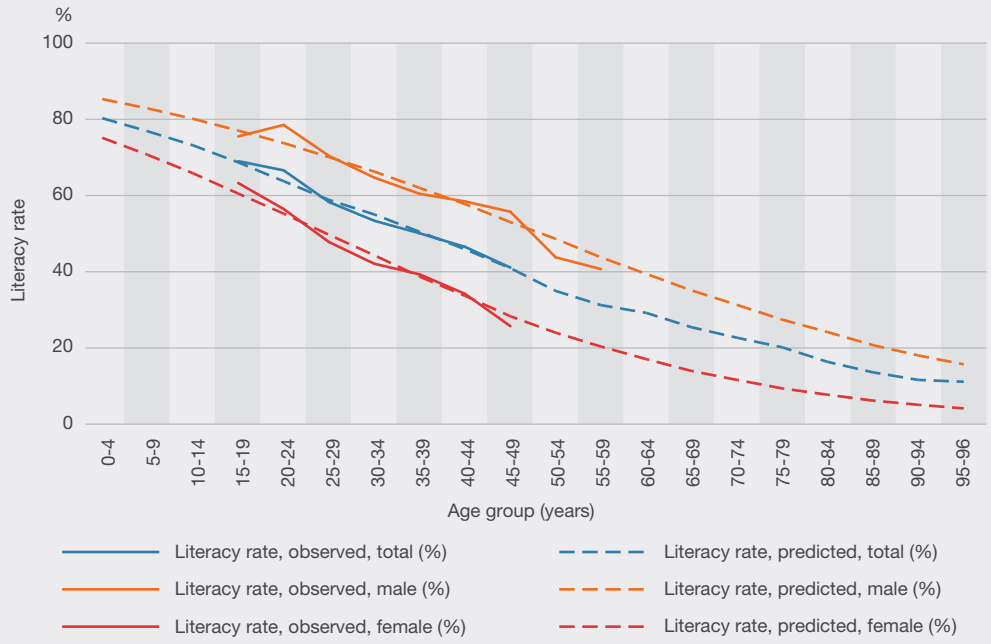
The most simplified version is the current dichotomous measure for literacy; it faces the most relevant challenge. For many years, the international definition of a literate person was someone "who can, with understanding, both read and write a short simple statement on his or her everyday life". This definition has long underpinned the UIS' regular Survey on Literacy which produces estimates of the literacy rates in most developing countries. These estimates, in practice, only distinguish between those who cannot read or write at all and the rest of the population. However, those judged to be literate relative to this definition can have vastly different levels of skills. Someone who can at best read and understand a simple statement about everyday life is arguably not sufficiently well-equipped to cope with the demands of modern-day living. Policy interventions are not only needed for those who are illiterate but also for those with weak literacy skills.

In order to address the needs of people with low literacy skills it is necessary to adopt a more nuanced definition of literacy which identifies a range of literacy skills and levels of competence. Being able to identify the characteristics not just of the illiterate population but also of those with weak skills will make it possible to better target resources to address their respective needs and increase literacy skills in general.

Self-assessment could be a simplified version of the type of DHS and MICS surveys that try to address the dearth of literacy assessments in developing countries by adding a simple test of reading skills to their survey modules. In DHS and MICS surveys, a sample of

**Figure 6.3 UIS literacy survey estimates**

Nigeria 2008 DHS: Observed and predicted literacy rate




*Source: Nigeria 2008 DHS.*

adult respondents, typically women and men between 15 and 49 years, are asked to read a card with a short, simple sentence in their language. The result is recorded as one of three options: i) cannot read at all; ii) able to read only parts of the sentence; or iii) able to read the whole sentence. The results of these tests are available for nearly all DHS and MICS surveys carried out in the last decade, including a large number of surveys in less-developed countries. The test results are more reliable than self-reported data on literacy and give at least some sense of the level of reading skills. On the other hand, these simple reading tests do not allow the measurement of literacy on a continuum, unlike the assessments mentioned earlier and are therefore only a partial improvement on traditional dichotomous literacy indicators.

### Skills surveys

Observed data could be either based on self-assessment or the administration of a skills survey that could have various alternatives and face alternative methodological decisions of the type described below:

- The number of skills domains;
- Testing the whole range of skills or limited to certain parts of the skills distribution;
- Whether the assessment will be conducted as an independent study or added to an existing study;
- Whether the assessment design will provide direct point estimates of skill distributions or support the generation of indirect, synthetic estimates; and
- Whether the assessment tool will be paper-and-pencil or computer-based.

### Synthetic estimates

An alternative is to do synthetic estimates based on observed available data *(see **Box 6.1**)*. The estimates could find various alternatives but consist, in a simplified version, of combining information regarding

### Box 6.1 Synthetic estimates to report for Indicator 4.6.1

A very simplified scheme would need:

- Data for either proportion of totals of persons at each education level by gender and age bands;
- Skills survey database to be prepped for modelling with as many countries as possible;
- The definition of a number of country-level factors according to available data that could be merged with the skills database described in the paragraph above; and
- Parameters to be estimated from the skills database to predict the values of the indicators.

*Source: UNESCO Institute for Statistics (UIS).*

the distribution of skills from countries that had administered skills surveys according to a defined set of characteristics of the population. This distribution of skills for those categories could be used to predict the levels (and, the indicator, once the minimum fixed levels of numeracy and literacy are defined) to establish the estimate for all countries.

These data would provide national and international users with a regular and current source of evidence, a prerequisite to maintaining policy focus and to adjusting policies and programmes. A large-scale rebasing of the model would be undertaken in 2031 when the next PIAAC collection cycle is undertaken.

Although some country-context parameters can be used, preliminary estimates show that some observable factors such as age, education and participation in some skills activities capture most of the variance. With this bridge, a country that has not administered a skills surveys but still has information about these parameters could have an estimate of the indicator. The degree of precision would vary according to the breadth of the individual information that could serve in the modelling phase.

An example of this type of modelling is the UIS literacy rate. Literacy rates for persons outside the age range with observed literacy rates are estimated using a logistic regression of literacy on age. As an example, see the 2008 DHS data for Nigeria in **Figure 6.3**. The survey collected information on literacy for women aged 15 to 49 years and men aged 15 to 59 years. The observed literacy rates are indicated by the solid lines and the results of the logistic regression are indicated by the dashed lines.

### 6.1.4 Reporting on the same scale

Once the definitions for the conceptual framework and levels of proficiency are sufficiently clear to allow options for countries to locate themselves in a continuum, the next step will be to develop an appropriate methodology for creating an internationally-comparable database to report on Indicator 4.6.1, given the use of different tools across countries. This means defining some criteria for linking that could be a combination of strategies or following a stepping-stone approach (as for Indicator 4.1.1).

The indicator requires the following inputs:

- Agreement on a proficiency framework that allows alternative levels of skills or functionality;
- Definition of the reference minimum global level;
- Definition of harmonisation (and linking) strategy that allows location of all efforts into a comparable metric; and
- A modelling strategy to produce an annual comprehensive set of literacy and numeracy estimates.

In order to produce estimates for reporting, there is a need for elaborating a guide and standards for countries that want to measure literacy, such as literacy modules to be added to household surveys and guidelines on the steps and standards for outputs. This type of global public good would facilitate not only a country's measurement but also commonalities in measurement. For this reason it is relevant to reach:

- An agreement of the questions in the self-assessment module;

## Box 6.2 Enhanced and shortened version of LAMP or mini-LAMP

UNESCO's LAMP assessment was developed to better respond to the needs of less-developed countries, while maintaining established proficiency scales. LAMP can be seen as a methodological endeavour to provide sound information, especially concerning the least-skilled in a population. It also shows the complexities of a diverse group of countries facing very different challenges in implementation. Through LAMP, the UIS has gained a unique perspective on the diversity of human literacy experience. Finally, it has also shown that the methodology, with the necessary adaptations, can be used across different cultures, languages and scripts.

Past experiences suggest the need for alternatives to a full LAMP assessment that would reduce the operational, technical and financial burden of fielding the assessment without compromising the ability to compare results across countries and over time. In this context, the UIS is taking a two-step approach to produce; i) a paper-and-pencil version; and ii) a device- or computer-based version of an enhanced and shortened version of LAMP, referred to as mini-LAMP.

Currently, the paper-and-pencil version of mini-LAMP has been produced and includes:

- Short literacy-relevant background questionnaire;
- Short cognitive modules;
- Administration guide;
- Translation and adaptation guide;
- Sampling guide;
- Scoring guide;
- Data capture and process guide; and
- Software and a data analytical guide.

To help countries with planning, the UIS will produce a national planning report template and memorandum of understanding to initiate discussions with interested countries

The device- or computer-based version of mini-LAMP is still under development.

*Source: UNESCO Institute for Statistics (UIS).*

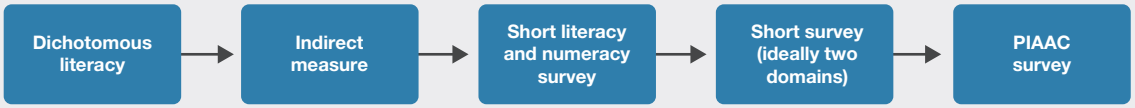
- A definition of the individual background question framing that would later serve as common parameters for synthetic estimates; and
- A short literacy numeracy module such as mini-LAMP *(see **Box 6.2**)* for those who want to do a shorter skills survey.

### 6.1.5 Laying out a strategy for measuring and reporting

Regarding the definition of literacy, GAML has recommended that the literacy and numeracy indicators be based on the framework of literacy and numeracy used in the OECD's PIAAC adult skills assessment programme. These definitions are precise enough to be measured and broad enough, with added elaboration at the foundational skills, to capture the entire range of skills encountered globally. Although the PIAAC assessment was only administered to 16- to 65-year-olds, the indicator covers 15-year-olds so information from PISA could also be used to report.

We propose a strategy for monitoring progress that offers countries a range of options according to their needs and possibilities. Countries on their way to achieving universal secondary education are encouraged to participate in the next round of the PIAAC data collection scheduled for 2021. The PIAAC design and processes are based upon 35 years of development and yield results that are valid, reliable, comparable and interpretable.

Figure 6.4 Country options from simple to complex




*Source: UNESCO Institute for Statistics (UIS).*

**Table 6.2 Cost of alternative options**

| Option | Estimated costs (US$) | Universe | Needs from countries |
|---|---|---|---|
| PIAAC | 2.5 million to 4 million[70] (paper-and-pencil and web-based) | ■ Country with experience in large-scale assessment and household survey<br>■ Strong technical capacity | Countries near achieving universal secondary education and have strong technical capacity |
| STEP | 500,000[71] (paper-and-pencil) | ■ Country with experience in large-scale assessment and household survey<br>■ Good technical capacity | Countries interested in literacy skills in working age population and have technical capacity |
| Short Literacy Survey[72] (SLS) | 200,000–400,000 (web-based) | ■ Country with experience in large-scale assessment and household survey | Developed countries that want more skills information beyond self-reporting and self-assessment but do not need a full range of skills estimates |
| Mini-LAMP[73] | 250,000–600,000[74] (paper-and-pencil)<br>160,000–300,000 (web-based) | ■ Country with experience in large-scale assessment and household surveys | Developing countries that want more skills information beyond self-reporting and self-assessment but do not need a full range of skills estimates |
| Literacy module (SLS or mini-LAMP) attached to DHS/MICS/ LFS | 150,000–200,000[75] | ■ Country with experience in large-scale assessment and household surveys | Countries that do not want to conduct a separate household survey for adult literacy but regularly conduct household surveys and want a snap-shot of targeted skills distribution |
| Synthetic estimation | Free-based on UIS methodology paper and set of guidelines on how to produce estimates | ■ Country with technical capacity | Countries that do not want to conduct another assessment but want to project skills using census data and existing assessment data to generate estimators to project future skills by sex and age group |

*Source: UNESCO Institute for Statistics (UIS).*

70 Based on communication with Scott Murray, the cost depends on sample size and cost of implementation within a given country.
71 Based on communication with the World Bank Group.
72 Since this is still under development, the estimate is based purely on speculation.
73 Estimate based on 1,500 cases with varied implementation costs from the UIS 4.6.1 option paper.
74 The cost is for a paper-and-pencil version and will be substantially smaller if it is attached to existing household surveys.
75 Estimate based on attaching the literacy module to a sub-sample of an existing household survey and cost of in-country training. No separate sampling cost as main sampling cost is borne by the surveyor.

For countries below this level of educational development, the current PIAAC design offers a limited information return on their investment. Moreover, the technical, operational and financial burdens imposed by PIAAC may be too great for some countries to bear, something that translates into a considerable risk of failure.

Each option comes with advantages and disadvantages, depending on what is valued more: skills coverage, reliability in generated estimates, accuracy in skills estimates and/or consistency in implementation. A combination of selected options could be chosen by countries and used to report on Indicator 4.6.1.

Hence, it is more realistic to consider a strategy that allows:

- Menu of options for countries that includes simpler to more complex alternatives to measure and report and allows countries to find their own model; and
- Use of estimates and projections to serve as a preliminary global picture of adult skills distribution.

In summary, each country has a choice on what works best for them. There are several options depending on socioeconomic development, as well as the technical and financial capacity of the country.

- A developed country that wants full skills distribution of its population could consider PIACC, which is technically-complex and expensive to implement.
- A developing country interested in understanding the literacy skills distribution of its productive population could consider STEP as it has comprehensive work-related background questions that provide precise skills distribution of the productive population.
- A country interested in only a targeted skills segment could consider SLS or mini-LAMP. Both of these short survey assessments consist of easier items which will provide better skills estimates for the country with a substantially-low literacy population.
- A country that wants a snapshot of its population's skills distribution could consider attaching a literacy module to an existing household survey. This will also reduce operating costs as the sampling cost has already been covered by the household survey.
- A country that has conducted literacy assessment in the past and does not want to conduct another round of adult literacy assessments could consider synthetic estimation. The UIS has developed a methodology paper on the way to produce a synthetic estimation based on basic characteristic variables such as sex, age group, years of schooling, etc. from a census. The relevant assessment data produced from past assessments, relevant characteristic variables and literacy-related questions can generate estimators to project skills distribution.

All options have their own advantages and shortcomings. Each country will need to identify what it considers most important to make the right choice.

## 6.2 PIAAC AND SDG MONITORING[76]

Indicator 4.6.1 is explicitly conceived as "a direct measure of the skill levels of youth and adults". Currently, the only comparable cross-country information regarding the proficiency of the adult population at the national level (i.e. persons aged 15 years or older) in literacy and numeracy based on the use of direct assessments is provided by PIAAC (OECD, 2013c and 2016a).[77] Three rounds of data collection have been undertaken in PIAAC in 2011-2012, 2013-2014 and 2017-2018. PIAAC is the third in a series of international assessments of adult literacy that has been implemented since the early 1990s that began with IALS over 1994-1998 (OECD/Statistics Canada, 2000) and the ALL over 1993-1997 (OECD/Statistics Canada, 2011). The PIAAC literacy assessment was designed to be linked

76 Written by William Thorn, Senior Analyst, OECD.
77 A full description of the methodology of the study is available in OECD (2016b).

with those used in IALS and ALL and the numeracy assessment is linked with that used in ALL.

STEP of the World Bank (Gaëlle et al., 2014) also includes a version of the PIAAC literacy assessment. However, the target population of STEP is, in most cases, the working age population (adults aged 15 to 64 years) in major urban centres whereas PIAAC covers adults of working age (16 to 65 years) who reside in the *national territory* of a participating country.

The OECD also manages an assessment of 15-year-old school students (PISA) which tests domains similar to those tested by PIAAC. While related conceptually, the assessments of literacy and numeracy in PISA are not psychometrically linked with assessments of literacy and numeracy in PIAAC, and results are not on the same scale. The relationship between the two studies is discussed in Chapter 6 of OECD (2016c).

To date, 38 countries have collected data as part of PIAAC. Results have been released for 33 countries participating in the first two rounds of data collection (2011-2012 and 2014-2015). Results from the third round will be released in 2019 *(see **Table 6.3**)*. A second cycle of the study using revised instruments is about to start with data collection planned for 2021-2022. The countries participating in PIAAC have been, in the vast majority, high-income countries. Some 15 middle- and low-income countries have participated in STEP.

Development of the second cycle of PIAAC started in early 2018. Data collection is planned to take place in 2021-2022, and the reporting of results at the end of 2023. At this point, it is expected that between 30 and 35 countries will participate.

Literacy is defined for the purposes of the PIAAC assessment as "understanding, evaluating, using and engaging with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential" (OECD, 2016c). Key to this definition is the fact that literacy is defined in terms of the reading of written texts and does not involve either the comprehension or production of spoken language or the production of text (writing).

Numeracy is defined as "the ability to access, use, interpret and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life". Numeracy is further defined in terms of the concept of "numerate behaviour" that involves managing a situation or solving a problem in a real context by responding to mathematical information and content represented in various ways (OECD, 2016b).

PIAAC results are reported on a 500-point scale in both literacy and numeracy, with higher scores representing higher proficiency.[78] To aid the interpretation of the scores, the scale has been divided into proficiency levels. The levels are defined as a score point range and are described in terms of the characteristics of the assessment tasks that a person who has a score in this range can successfully complete with a reasonable chance of success. Six levels are defined, ranging from less than Level 1 (the lowest) to Level 5 (the highest) in both literacy and numeracy. The cut-points are presented in **Table 6.4**.

The features of Level 1 tasks in literacy and numeracy are described in **Table 6.5** by way of example.[79]

The mean literacy score and the proportion of the population that has achieved the different proficiency levels for 29 countries in Round 1 and 2 of PIAAC (Cyprus and the Russian Federation are not included) are presented in **Figure 6.5**. As can be seen, there is a close correlation between the average score and the distribution of the population across the proficiency levels. Countries with higher mean scores have smaller proportions of their population in the lowest two proficiency levels.

78 The mean score (OECD countries) is 268 score points in literacy and 263 score points in numeracy. The standard deviation on both scales is slightly less than 50 score points.
79 See OECD (2016c) for the descriptors for other levels.

**Table 6.3 Countries participating in PIAAC and STEP**

| PIAAC Round 1 (2011-2012) | PIAAC Round 2 (2013-2014) | PIAAC Round 3 (2017-2018) | PIAAC 2nd Cycle (2021-2022) | STEP* (2012-2017) |
|---|---|---|---|---|
| Australia | Chile | Ecuador | Australia | Armenia |
| Austria | Greece | Hungary | Austria | Azerbaijan |
| Canada | Israel | Kazakhstan | Canada | Bolivia |
| Czechia | Lithuania | Mexico | Chile | Colombia |
| Cyprus | New Zealand | Peru | Croatia | Ghana |
| Denmark | Portugal | United States | Czechia | Kenya |
| England | Slovenia | | Denmark | Georgia |
| Estonia | Singapore | | England | Laos |
| Finland | Turkey | | Estonia | FYR of Macedonia |
| Flanders | | | Finland | Serbia |
| France | | | Flanders | Sri Lanka |
| Germany | | | France | Ukraine |
| Ireland | | | Germany | Viet Nam |
| Italy | | | Iceland | Yunnan (China) |
| Japan | | | Ireland | |
| Korea | | | Israel | |
| Netherlands | | | Italy | |
| Northern Ireland | | | Japan | |
| Norway | | | Korea | |
| Poland | | | Latvia | |
| Russia | | | Lithuania | |
| Slovakia | | | Netherlands | |
| Spain | | | New Zealand | |
| Sweden | | | Norway | |
| United States | | | Poland | |
| | | | Russia | |
| | | | Slovakia | |
| | | | Singapore | |
| | | | Spain | |
| | | | Sweden | |
| | | | Switzerland | |
| | | | United States | |

**Note:** *Population in urban centres.
*Sources: PIAAC and OECD.*

It is important to note that the purpose of the proficiency levels in PIAAC is descriptive (OECD, 2013a). They are intended to facilitate the interpretation and communication of the results by describing the characteristics and features of the assessment tasks that a person with a particular proficiency score can typically complete successfully. They have no normative purpose and should not be interpreted as representing performance standards or benchmarks. In particular, the cut-points between levels are related to particular features of the scales and there are no natural breaking points along the scales that could be used to separate different levels of proficiency. Other cut-points, different numbers of levels, and other bandwidths could have been selected to define the proficiency levels with equal

**Table 6.4 PIAAC literacy and numeracy levels, score point ranges**

| Level | Score point range |
|---|---|
| Less than 1 | 0-175 |
| 1 | 176-225 |
| 2 | 226-275 |
| 3 | 276-325 |
| 4 | 326-375 |
| 5 | 376-500 |

*Source: OECD, 2016b.*

**Table 6.5 Descriptors of Level 1 tasks in literacy and numeracy**

| Literacy |
|---|
| Most of the tasks at this level require the respondent to read relatively short digital or print continuous, non-continuous, or mixed texts to locate a single piece of information that is identical to or synonymous with the information given in the question or directive. Some tasks, such as those involving non-continuous texts, may require the respondent to enter personal information onto a document. Little, if any, competing information is present. Some tasks may require simple cycling through more than one piece of information. Knowledge and skill in recognising basic vocabulary determining the meaning of sentences, and reading paragraphs of text is expected. |
| **Numeracy** |
| Tasks at this level require the respondent to carry out basic mathematical processes in common, concrete contexts where the mathematical content is explicit with little text and minimal distractors. Tasks usually require one-step or simple processes involving counting; sorting; performing basic arithmetic operations; understanding simple percentages such as 50%; and locating and identifying elements of simple or common graphical or spatial representations. |

*Source: OECD, 2016c.*

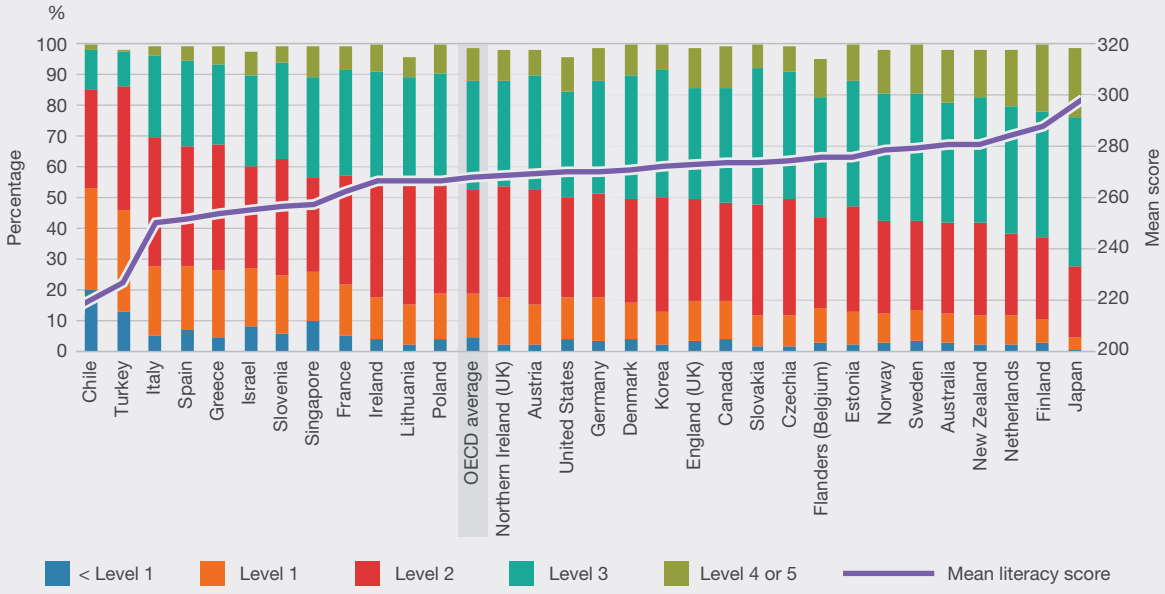
justification.[80] It is also important to note that the proficiency levels in literacy and in numeracy should not be seen as equivalent in any sense. The scales do not measure the same constructs and literacy and numeracy items located at the same nominal point on their respective scales cannot be said to be of equivalent difficulty. Thus, it is not meaningful to compare the proportion of the population in Level 1 in literacy with that in numeracy, for example.

Normative interpretations of the proficiency levels in adult literacy surveys have been proposed. In IALS, a predecessor of PIAAC, the claim was made that Level 3 in literacy could be "considered a suitable minimum for coping with the demands of everyday life and work in a complex, advanced society" (OECD and Statistics Canada, 2000). The empirical basis for this claim was weak. Treating Level 3 as the suitable minimum level of performance in literacy had little face validity. Almost all the countries in IALS, most of which were "advanced" countries, had at least 40% of their population with proficiency below Level 3 and many had over 50%. Defining lower levels on the PIAAC scales as "minimum suitable levels of proficiency" faces similar problems. A recent report looking at the population scoring at Level 1 or below in PIAAC in literacy and in numeracy concluded that: "Low proficiency adults are not sharply differentiated from the rest of the adult population in terms of sociodemographic characteristics considered either across or within countries" (Grotlüschen, et al., 2016). For example, while the probability of a person with literacy proficiency at Level 1 or below being employed was lower than for the rest of the population, most adults in this group were, nevertheless, employed.

A good discussion of the complexities inherent in any attempt to define thresholds that represent minimum

80 For a good discussion of this issue see OECD (2006) which refers to PISA but which is equally relevant to PIAAC. "It is important to understand that the literacy skills measured in PISA must be considered as continua: there are no natural breaking points to mark borderlines between stages along these continua. Dividing each of these continua into levels, though useful for communication about students' development, is essentially arbitrary. Like the definition of units on, for example, a scale of length, there is no fundamental difference between 1 metre and 1.5 metres – it is a matter of degree. It is useful, however, to define stages, or levels along the continua, because this enables communication about the proficiency of students in terms other than numbers."

**Figure 6.5 Mean literacy score and percentage of the population by proficiency level**




**Note:** Percentages may not add up to 100% due to a group of the population who did not answer the background questionnaire or take the assessment for language-related reasons.
*Source: Survey of Adult Skills (PIAAC), 2012, 2015.*

desirable, sufficient or adequate levels of literacy (and by extension numeracy) can be found in Maddox and Esposito (2011). This describes the issues that must be faced in establishing minimum levels, such as the arbitrariness of any cut-point, as well as the complexity of their interpretation.

Indicator 4.6.1 reports the proportion of the adult population (16-65 years of age) scoring at Level 2 or above (i.e. who have a score equal or greater than 226) on the literacy and numeracy scales respectively for the countries that have participated in PIAAC. As is clear from the above, this figure should not be interpreted as the proportion of the population who possess skills above a "benchmark of basic knowledge" or as an estimate of the proportion of the population possessing an "adequate" or "sufficient" level of proficiency in either literacy or in numeracy. At most, it can be interpreted as offering an indication of the proportion of the population that has the capacity to successfully complete reading tasks that involve locating single pieces of information in short texts or numeracy tasks that involve simple mathematical processes.

In order to gain a comprehensive and nuanced picture of the literacy and numeracy proficiency of the adult population in the countries covered by PIAAC (and other countries where equivalent data exist), it is important to look beyond single indicators such as 4.6.1. Interested readers are referred to the reports of Round 1 (OECD, 2013c) and Round 2 (OECD, 2016a) of PIAAC for a detailed presentation of the results in the countries participating in PIAAC.

Looking forward, data for five additional countries in the first cycle of PIAAC will be released in 2019. Data from the second cycle of PIAAC will be released in late 2023. This will provide an opportunity to look at change in the proficiency in literacy and numeracy of the working age population in most participating countries between 2011-2012 or 2013-2014 and 2021-2022. In addition, for countries that participated in IALS, ALL or both, comparisons will be able to be made over longer periods of time. As PIAAC is planned on a ten-year cycle, this is likely to be the only observation of the literacy and numeracy proficiency of the working-age population that will

be available for these countries during the reporting period for the SDGs (2016-2030). It is possible that, as in the first cycle, further rounds of PIAAC covering additional countries may take place as part of the second cycle.

## 6.3 USING THE STEP HOUSEHOLD SURVEY TO INFORM INDICATOR 4.6.1[81]

### 6.3.1 The learning crisis and the power of adult learning

Adult literacy helps individuals work productively, live healthy lifestyles and improve life satisfaction. In this way, those empowered can better contribute to economic prosperity and social progress. This is an important reason why adult literacy is a key component of the SDGs. Despite the considerable benefits of literacy, many adults in low and middle-income countries are still functionally illiterate, which is in stark contrast to the rapid increase in the educational attainment these countries have achieved over the last decades (World Bank, 2018). There is clearly a need to better understand the nature of the skills shortages and the population sub-groups with a large proportion of illiterate adults, with a view to identifying appropriate policy measures and instructional responses to address the learning crisis. The World Bank's STEP Household Survey is designed to help fill this gap in knowledge.

### 6.3.2 Results from the STEP household survey

The STEP household survey is an international skills assessment programme that sheds light on adult literacy and socio-emotional skills in low- and middle-income countries (World Bank, 2014). STEP is one of the very few international assessments specifically designed to measure adult skills in developing countries, and the only assessment that provides literacy measures that can be linked to the OECD's PIAAC proficiency scale *(see Section 6.2).* STEP's Waves 1 to 3 have already been administered in 15 countries between 2011 and 2016, including ten countries that provide results of the full literacy assessment.[82] STEP's literacy assessment is designed to measure adult's capacity to understand, evaluate, use and engage with written texts, while the socio-emotional skills assessment, partly based on items from the Big Five Inventory (BFI),[83] aims at capturing adult's diverse psycho-social characteristics and behaviours.

STEP's literacy assessment allows identification of the levels and distributions of skills as well as their correlates with individual background and behavioural outcome measures. STEP adopts PIAAC's six proficiency levels: Below Level 1, Level 1, Level 2, Level 3, Level 4 and Level 5. Given STEP's focus on the lower levels of the PIAAC literacy scale, approximately 89% of the items fall between Below Level 1 and Level 3, although there are also items (11%) that cover Level 4 and 5 (ETS, 2014). The assessment focuses on reading literacy and does not include numeracy or other cognitive domains such as problem-solving in a technology-rich environment. While STEP's background questionnaire (which includes the socio-emotional skills assessment) is delivered using interviewers, the literacy assessment component is self-administered by adult test-takers using paper and pencil.

**Figure 6.6** represents the proportion of adults (aged 15 to 64) who scored at or above the minimum literacy proficiency threshold level, which is equivalent to PIAAC proficiency Level 1,[84] which corresponds to

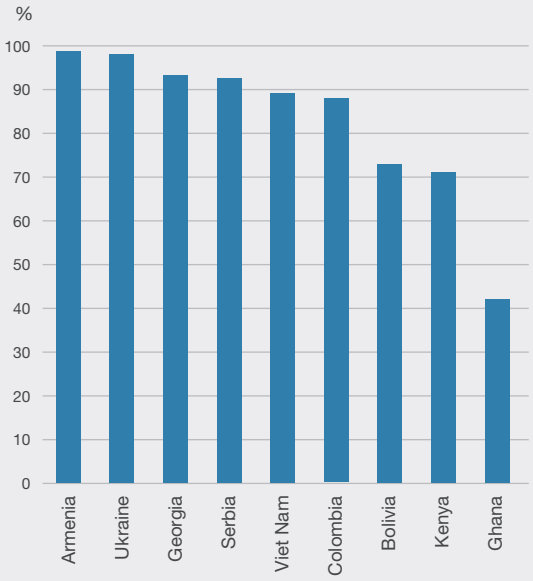
81 Written by Koji Mijamoto, Senior Economist, Education Global Practice, World Bank Group.

82 STEP's micro-data and related reports can be found in the World Bank's micro-data library: http://microdata.worldbank.org/index.php/catalog/step

83 The Big Five Inventory (BFI) is a self-report inventory designed to measure the Big Five dimensions. BFI use short phrases with relatively accessible vocabulary. STEP used an adapted version of the BFI items.

84 Adults who are at or above the PIAAC Proficiency Level 1 are considered capable of performing tasks that require the respondent to read relatively short print continuous, non-continuous or mixed texts to locate a single piece of information which is identical to or synonymous with the information given in the question or directive. Some tasks may require the respondent to enter personal information into a document, in the case of some non-continuous texts. Little, if any, competing information is present. Some tasks may require simple cycling through more than one piece of information. Knowledge and skill in recognising basic vocabulary, evaluating the meaning of sentences, and reading of a paragraph of text is expected (ETS, 2014).Note that the World Development Report uses PIAAC Proficiency Level 2 as the "minimal level of foundational literacy" or "low proficiency" (World Bank, 2018).

**Figure 6.6 Percentage of working-age population who are at or above the minimum literacy threshold, 2011-2016**




**Note:** Data are based on the latest availability from STEP Skills Measurement Program. STEP is representative of urban populations, aged 15 to 64. Those who are considered at or above the minimum literacy threshold have demonstrated literacy proficiency at Level 1 or higher.
*Source: STEP Skills Measurement Program, 2011-2016 (http://microdata.worldbank.org/index.php/catalog/step/about).*

a level that captures the adult's capacity to read short texts to locate a single piece of information. The figure suggests considerable cross-country differences: some low-income countries such as Kenya and Ghana have a large proportion of adults who cannot demonstrate the minimum literacy proficiency, while a number of middle-income countries including Armenia, Ukraine, Serbia and Georgia have a relatively large proportion of adults who are at or above the minimum literacy threshold.

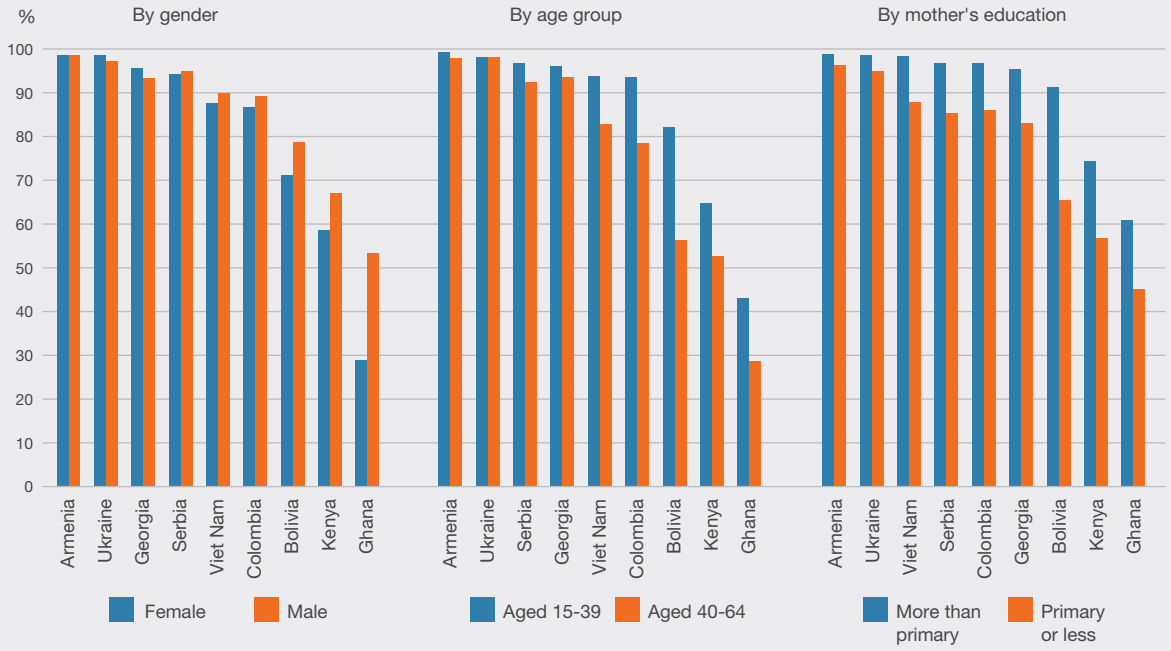
Are there particular population sub-groups with a larger proportion of adults who have attained the minimum literacy proficiency threshold? Panels A, B and C in **Figure 6.7** present the proportion of adults who scored at or above the minimum literacy proficiency threshold by sex, age and mother's educational attainment. The panels suggest that in most countries, adults who are male, young and with mothers who have completed more than primary schooling are more likely to score at or above the minimum literacy threshold than otherwise. The panels also show that for those countries with a lower proportion of adults scoring at or above the minimum literacy threshold, there is a larger disparity in literacy across sex, age and mother's education. For instance, Panel A shows that in Bolivia, Kenya and Ghana, men are considerably more likely to score at or above the minimum literacy threshold than women. Panels B and C also show that, for these three countries, a much larger proportion of adults who are younger and with mothers with a degree higher than the primary school level score at or above the minimum literacy threshold. These results suggest that sex, age and parental education may play important roles in addressing education policies and practices related to addressing adult illiteracy.

### 6.3.3 How STEP indicators can help countries work towards SDG 4

Results from the World Bank's STEP household survey demonstrate the powers of mobilising large-scale skills assessments in highlighting the nature and intensity of the skills shortages and the population sub-groups that demand urgent attention. These surveys offer policymakers and practitioners' background information to explore the development of strategies and implementation plans to address the learning crisis. Moreover, internationally-comparable skills assessments allow countries struggling to improve adult literacy to not only understand the skills shortages vis-à-vis other countries, but also the experiences of successful reformers. In this way, international assessments such as STEP can provide valuable inputs to countries striving to achieve Indicator 4.6.1.

Low- and middle-income countries interested in assessing skills may choose to administer national skills assessments or join international initiatives such as STEP. By joining STEP, countries have the possibility of measuring adult literacy on a PIAAC

**Figure 6.7 Percentage of working-age population who are at or above the minimum literacy threshold, by sex, age and mother's education, 2011-2016**




**Notes:** Data are based on the latest availability from STEP Skills Measurement Program. STEP is representative of the urban population, aged 15 to 64. Those who are considered at or above the minimum literacy threshold have demonstrated literacy proficiency at Level 1 or higher. Those at this proficiency levels can execute tasks that require the respondent to read relatively short print continuous, non-continuous or mixed texts to locate a single piece of information which is identical to or synonymous with the information given in the question or directive.
*Source: STEP Skills Measurement Program, 2011-2016 (http://microdata.worldbank.org/index.php/catalog/step/about).*

proficiency scale, thereby allowing comparison of results with a range of low-, middle- and high-income countries that have participated in PIAAC or STEP. Moreover, for countries that have limited experience and technical capacity to administer complex surveys, participation in STEP would allow delivery of a skills assessment that complies with the technical standards and protocols that the World Bank has developed in collaboration with the OECD and the Educational Testing Service (ETS).

## 6.4 DEVELOPING EVALUATION CAPACITY AND ACTION RESEARCH IN AFRICA[85]

This section reports on an evaluation capacity development model in the context of the emerging, results-oriented culture in Africa. It is not meant to be prescriptive, but focuses on the relevant elements of this process and the experience of *Recherche-action sur la mesure des apprentissages des bénéficiaires des programmes d'alphabétisation* (Action Research on Measuring Literacy Programme Participants' Learning Outcomes) (RAMAA).

RAMAA was initiated by the UIL at the request of certain African countries. It focuses on the field of non-formal education and aims to provide policymakers and development partners with reliable and comparable data adapted to the quality of literacy programmes.[86]

Its objective is to assist countries in setting up a system for monitoring and evaluating the quality of

85 This section is based on Bolly, 2018.

86 Literacy programmes refer to organized learning arrangements that target young people and adults who are illiterate or have low literacy skills.

literacy provision. Starting from the development of a standardised methodological framework, it aims to provide a better understanding of the proficiency levels of reading, writing, mathematics and problem-solving skills acquired by youth and adults aged 15 years and older who are participating in literacy programmes. This is supported by an assessment of the determinants of quality that specify the contextual variables that explain the different outcomes between participants and countries.[87] This information mechanism is designed to effectively and regularly guide literacy policies.

In short, RAMAA focuses on the participants of literacy programmes as part of a population. Its reports results at the global level, rather than the individual level, in order to assess the results of the literacy sector in the RAMAA countries and their evolution over time. In terms of impact, this review of literacy outcomes by RAMAA will test whether literacy programmes provide participants with a common core of basic skills, thus making literacy one of the sub-sectors of education that contributes to a more inclusive and just society. Taken in terms of quality of education, as well as skills and learning outcomes, RAMAA is contributing to Indicator 4.6.1 and broader aims of the SDGs.

RAMAA is designed in the spirit of action research; the project itself is under development. The experience of the first phase (2011-2014), which started with five countries (Burkina Faso, Mali, Morocco, Niger and Senegal) led to the expansion to seven other countries (Benin, Côte d'Ivoire, Cameroon, Central African Republic, Chad, Democratic Republic of Congo and Togo) in the current phase (2016-2020). It was decisive for at least three reasons:[88] i) three out of five countries demonstrated strong political commitment that has resulted in substantial financial and technical mobilisation; ii) the mobilisation of national experts with multidisciplinary profiles led to a collective and successful learning dynamic; and iii) the results have helped to reshape national literacy strategies in some countries (Morocco, Niger and Senegal), as well as the potential development of master's degrees in education sciences at national universities (e.g. Senegal).

We have chosen to focus on modalities for the development of evaluative capacity because of its poorly documented nature, unlike the technical aspects that are widely debated in many scientific papers (Varone, 2007). The evaluative capacity development model in RAMAA is original and differs from the top-down, vertical logic that characterises most international surveys. Conducted horizontally in the context of Action Research, this approach provides a solid foundation for the effective implementation of a results-oriented culture.

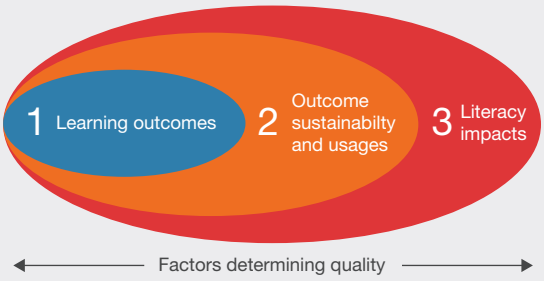
### 6.4.1 Methodological framework of RAMAA

The quality of the youth and adult literacy programmes in RAMAA is captured at three distinct levels: i) the level of learning of beneficiaries upon beginning and end of the literacy programmes; ii) the sustainability of learning over time and space; and iii) the impacts of literacy *(see **Figure 6.8**)*. In the current phase, the 12 RAMAA countries are engaged in the production of standardised measurement tools in accordance with the first level of analysis.

The organization of the survey is spread over a period of five years (2016-2020) and sub-divided into four phases: i) the first year focuses on the consolidation of the partnership framework; ii) the next two years are devoted to the updating of contextual data, the adjustment of measurement instruments and collection methods;[89] iii) the fourth year gives way to the pilot testing of measuring instruments and collection procedures; and iv) the fifth year gives rise to the execution of the assessment itself.

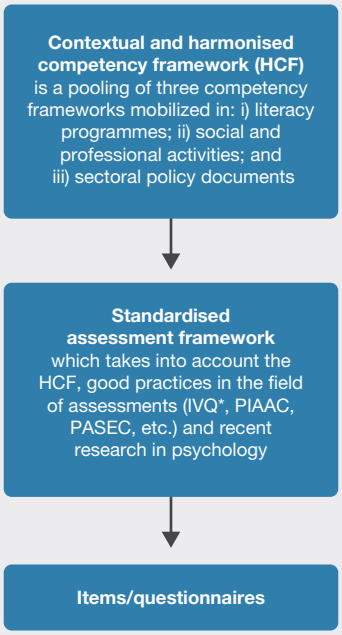
87 The determinants of quality are essential in the sense that the production function of literacy programmes is not uniform but variable in terms of populations, operators and approaches, among others.
88 See Bolly and Jonas, 2015.
89 This includes the updating of the harmonised competency framework, the development of the assessment framework as well as the elaboration of the items/questionnaires.

**Figure 6.8 Different levels of analysis of RAMAA learning outcomes**




*Source: Diagram developed by Sobhi Tawil, UNESCO.*

**Figure 6.9 Methodology for the development of learning measurement tests towards standardised measurement tools**




**Note:** * IVQ is the Survey on Information Exchange and Daily Life
*Source: UIL.*

Currently, RAMAA is in the adjustment stage of the measurement tools, including the assessment framework that specifies the skills to be assessed, the content standards for these skills and the psychometric method for determining proficiency levels.[90]

The originality of the RAMAA assessment framework is that it relies initially on fine contextual analysis of the common frames of reference associated with the literacy provision of the RAMAA countries, which, in turn, will allow it to incorporate good practices in literacy assessment *(see **Figure 6.9**)*.

This first descriptive construct, which we call the RAMAA Harmonised Competency Framework (HCF), aims to identify the profile in terms of literacy skills of a so-called "literate" person in the context of the RAMAA countries, by pooling:

- Competency frameworks available in the literacy programmes of the countries concerned;
- Competency frameworks mobilised in social and professional activities; and
- The competency framework describing the ideal profile of the literate citizen as reflected in sectoral social policy documents (education, employment, health, social development, etc.).

A second RAMAA specificity relates to the consideration of writing skills, which are poorly reflected in international surveys, due in part to the difficulty of comparing performances in different writing systems (Jeantheau, 2015). The first phase of RAMAA included twelve national languages. However, the heterogeneity of the exercises developed during this phase made the data processing more complex. In terms of remediation, this phase of RAMAA will standardise and limit the collection of information. It will largely rely on the methodological approach of the survey on information exchange and daily life.

The last RAMAA specificity lies in the measurement, in declarative form, of the socio-educational and professional skills common to the context of RAMAA countries. In the first phase of the project, they

90 See Mally (2018) for details.

focused on four areas, namely health/well-being, citizenship, environment and work.

Proficiency levels will not be fixed in advance but according to an empirical data mining procedure. It is the success of the participants that will allow us to assess the levels of control defined not in a dichotomy but on a continuum (see procedure adopted by PIAAC). Thus, from the answers given by the participants to the test items, a performance scale will be developed using an item response model. As per the theory related to this type of model, the scores of the respondents as well as the level of difficulty of the items are measured on the same scale. This makes it possible to build groups of levels and to associate them with sets of items of increasing difficulty (Rocher, 2015).

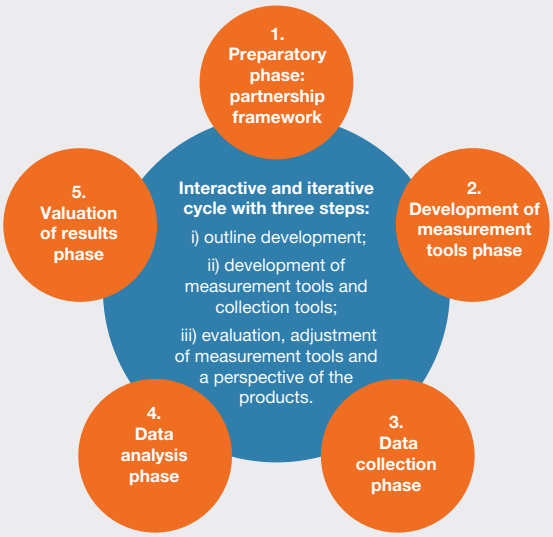
### 6.4.2 RAMAA model of developing assessment capacities for anchorage in a results-oriented culture

By "assessment capacities" in RAMAA, we mean the ability to put in place a sustainable mechanism for monitoring and assessing the quality of the literacy programmes. The viability of such an enterprise is based on interdependent guiding principles, which include the following:

- Institutional: a firm political will which translates into the signing of a memorandum of understanding and a technical (making use of a set of national skills) as well as financial commitment (taking charge of national activities) of the countries in RAMAA.
- Strategic/organizational: action research that favours a dynamic co-construction based on the active participation of national teams at all stages of the RAMAA programme. The goal is to promote a results-oriented culture.

Overall, skills development should not be seen as a goal in itself, but as a process that takes time to make a real impact. The investment will yield future returns. This co-construction operates:

**Figure 6.10 RAMAA model for assessment capacity development**




*Source: UIL.*

i) Throughout the RAMAA implementation by pooling experiences in the form of an interactive and iterative cycle with three stages *(see **Figure 6.10**):*

- Development of the frameworks (guidelines);
- Development of measurement tools and collection tools; and
- Evaluation and adjustment of measurement tools and country perspective (national reporting). The multi-cultural professionalism of national experts, external experts and UNESCO are a source of mutual enrichment in the sense stressed by Courtois (2013).

ii) South-South cooperation boils down to pooling expertise and strengthening inter-country reciprocal links.

### 6.4.3 Conclusion

RAMAA offers a standardised measure of reading, writing, mathematics and problem-solving skills targeted at youth and adults (aged 15 years and older) who benefit from literacy programmes. The data and analyses produced by RAMAA will be made available to policymakers in the countries concerned, as well as the educational community of researchers and civil society, with the aim of contributing to the debate on the quality of education and the governance of literacy programmes. While emphasising the assessment of functional writing skills and generic skills, RAMAA also can contribute to monitoring progress towards SDG Target 4.6 and, in particular, Indicator 4.6.1. The collaboration of a wide range of national and international experts and the application of rigorous standardised procedures offer guarantees of reliability and sustainability.

The real challenge today is funding. The inclusion of assessment in national budgets remains modest and has slowed the implementation of RAMAA. Achieving the goal of a genuine institutional anchoring of a national assessment policy requires stronger political and financial support. The latter is a fundamental necessity for achieving SDG Target 4.6.

# 7. Supporting countries to produce learning data for Indicator 4.1.1

Learning is key for the personal, social and economic development of countries. To know how much children and youth are learning, it is imperative to have learning assessments in place. Assessments allow countries to monitor and support learning by informing educational policy and practice. These assessments can also be used to inform the SDGs and therefore contribute to monitoring learning globally.

Countries may face tough decisions when implementing learning assessments. For instance, a country may need to decide if it is better to start by assessing mathematics or reading, or conduct a full assessment at Grade 3 or the end of primary education. Scarce resources and limited local capacity may force countries to pick one over the other or to make a long-term plan stating which assessment will be implemented first.

Countries need to make strategic decisions regarding what type of assessment to implement (national, regional and/or cross-national). If they decide to do a national assessment, they will need to make additional decisions regarding what to measure, who to measure, when to measure, among others. In making these decisions, it is important that countries take into account their national education goals, priorities and resources. It is also important that they take into account the requirements to inform the SDGs. By doing so, they will maximise the use and potential benefit of their data.

Technical and financial assistance is also needed so that countries can produce the data to inform SDG 4. For instance, they may need to strengthen sampling procedures to ensure that assessment results are representative at the national level and not of urban schools only. Offering hands-on training to the assessment team has been a successful strategy to develop local capacity. However, to conduct a reliable and valid assessment, capacity development and standards concerning best practices and reporting must be in place.

This chapter presents guidelines for countries aiming to implement an assessment to monitor learning at the national level and to inform SDG 4. The first section reviews current sources of information and how they can be used to report on SDG indicators. Section 7.2 explores the options and challenges facing countries in implementing a learning assessment. Section 7.3 explores ideas about how to share learning assessment information with different stakeholders.

## 7.1 HOW LEARNING ASSESSMENTS COULD INFORM SDG 4 INDICATORS

The SDG 4-Education 2030 Agenda presents national and international education stakeholders with two important measurement challenges: learning outcomes and educational equality. Equity focuses on the need to take into account the many aspects related to those who have been left behind. The SDG Agenda includes equity-specific goals (Goal 5 on gender equity and Goal 10 on reducing inequalities). Until recently, global monitoring of inequalities in education and other sectors has mainly captured differences by sex. The SDGs have a broader scope including wealth, location, ethnicity, language and disability, as well as inherent variability (between

## Box 7.1 Raising the floor of learning levels: Equitable improvement starts with the tail

Learning levels among the majority of children in developing countries often do not meet the expectations of national curricula, nor basic levels of competence tested in citizen-led assessments (e.g. ASER, Uwezo). The learning crisis has been well documented, in addition to the systemic failures. This may explain the prevalence of poor learning outcomes and remains a key area of study.

While only a few pupils in developing countries reach learning levels comparable to OECD norms, de facto exclusion from minimally-acceptable learning competences represents both a failure of education systems and a global "equity crisis". Poor learning among children, especially where it is a result of poor quality education, is inequitable as it contributes to massive global (North-South) inequality. It also contributes to failures to develop and realise the talents of all pupils. This latter form of inequity is linked to absolute notions of right or entitlement; or in Sen's terms (reference), to the right to opportunities to develop valuable human "capabilities" and "functioning", in which education plays a key role. The right to education, enshrined in the UN Declaration of Human Rights, is founded on the development of such capabilities, not simply on schooling. The SDGs represent an opportunity to focus on learning and its distribution. These goals, which replace the MDGs, focus primarily on learning, not just a minimum proficiency approach (increasing the percentage of children reaching a minimum level of proficiency) and inequality, which is consistent with the empirical patterns and themes documented in this note.

Educational inequalities in developing countries are typically high (higher than income inequalities in some cases), while average performance levels remain low (striking examples include South Africa and India). OECD evidence suggests that educationally high-performing countries tend to also have lower levels of inequality, i.e. higher average learning levels are associated with lower inequality in learning levels. Understanding how to reduce inequalities, while simultaneously raising learning outcomes, remains an important question for education stakeholders.

*Source: Crouch and Rolleston, 2017.*

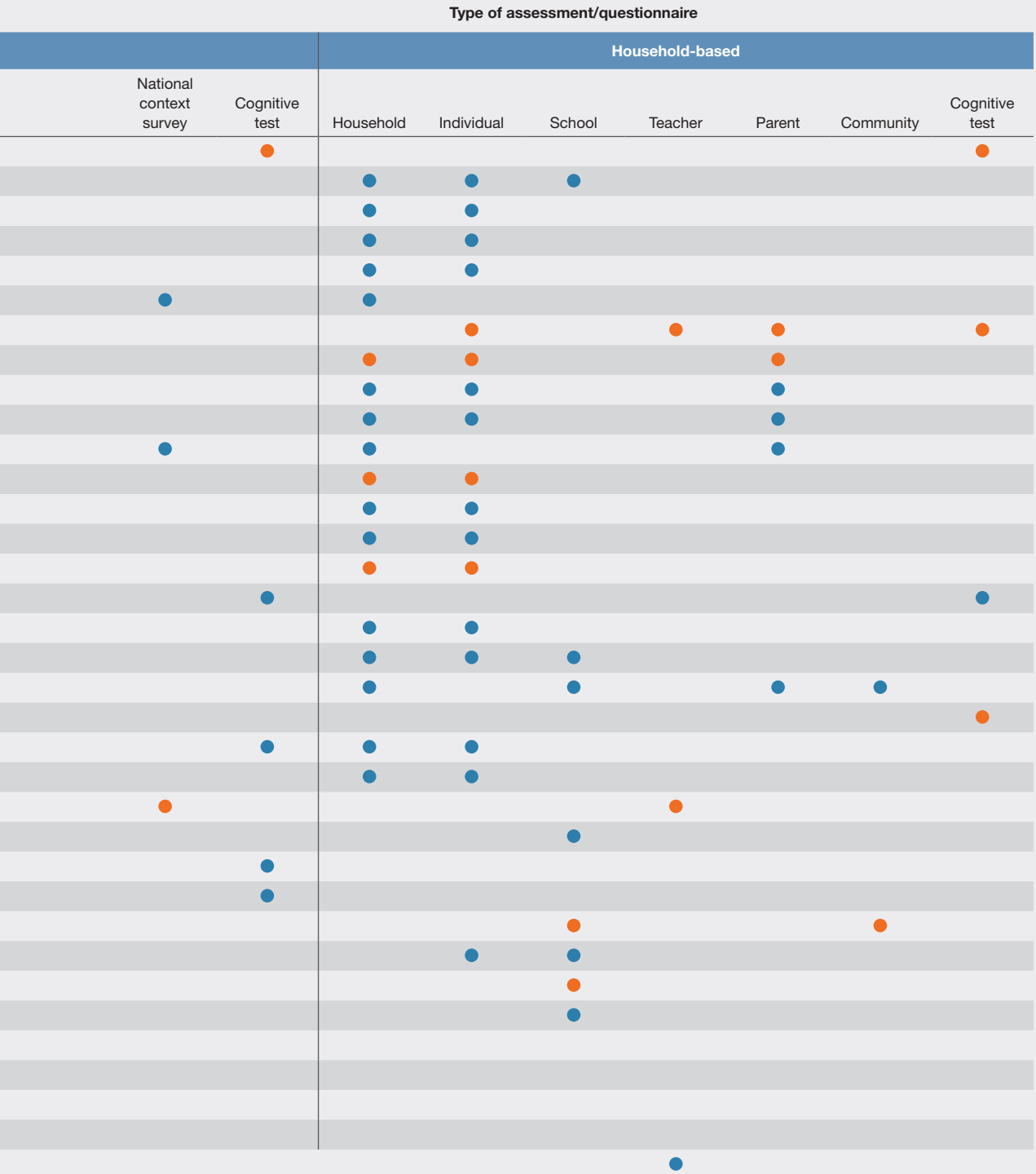
schools and students), as this is the biggest source of inequality in most countries *(see **Box 7.1**)*.

Though tremendous strides have been made in increasing access to schooling, marginalised populations, such as the poor, disabled, displaced or nomadic populations, are frequently under-represented in schools. Similarly, even when marginalised or disadvantaged populations are able to attend school, they often attend under-resourced and poor-quality schools with lower student proficiency rates. Continuing efforts are needed to ensure that all children are attending quality schooling. SDG 4 is designed specifically to highlight the need to support all children's access to and success in school. Just as worryingly, but also holding great potential, is the fact that in some countries there is great variability of results *among* the poor or *among* the vulnerable. This suggests poor quality control as a factor of "pure" or inherent inequality. But it also holds great potential because it suggests (using evidence from the countries themselves) that it is possible to improve results even for the poor or more vulnerable, as many schools attended by the lower-income segments of society sometimes perform as well as schools attended by those from higher-income segments.

The focus on learning is more demanding, but also more meaningful, than a focus only on access to schooling, because there is far more inequality in learning around the world than there is in access to schooling. As an example, Crouch and Gustafsson (2018) found for reading skills that the inequality in learning outcomes is 170% greater than the inequality in access to secondary education and 43% greater than the inequality in access to tertiary education. The SDGs still do not cover what one might call "pure" or "total" inequality, that is, the total dispersion in scores, due to factors such as income and region, but also importantly due to a lack of quality assurance and standards.

## Figure 7.1 Map of SDG 4 global and thematic indicators in learning assessment questionnaires


*Source: UNESCO Institute for Statistics (UIS), http://gaml.uis.unesco.org/dashboard*


Type of assessment/questionnaire
Household-based
National context survey
Cognitive test
Household
Individual
School
Teacher
Parent
Community
Cognitive test

### 7.1.1 What information can learning assessments collect?

Learning assessments can play a pivotal role in informing efforts to achieve SDG 4. This section reviews the feasibility of using learning data to report on SDG indicators for cognitive and other purposes. Due to the wide variety of indicators included in SDG 4, both household-based surveys and school-based assessments collect background information that put the data in context. By covering children and young people in school and out, household-based surveys provide information on households and enabling environments. School-based assessments provide system-level information on the classroom and school environment.

Together, household-based surveys and school-based assessments help to present a snapshot of how children and youth around the world are learning. These evaluations provide information on more than simply cognitive outcomes. They include information on context and factors, through student, family, teacher and school background questionnaires, which could affect those outcomes. Data are disaggregated by criteria such as sex, age, location (rural/urban), socioeconomic status, language spoken at home, ethnic group, immigration status, disability, etc. In addition, there is information on household characteristics associated with out-of-school populations. For example, these surveys can capture information on the education levels of parents, health, nutrition, disability and family support, including attitudes about school and expectations for the family's children. Data collected through household surveys can be used to estimate demand for and barriers to school attendance.

In general, background questionnaires from school-based assessments include principals (school heads, head teachers), schools, teachers, ICT coordinators, students, homes, curriculum and national context surveys, while household-based assessments include parents (caregivers), schools, teachers, individuals (children, adults) and communities.

**Figure 7.1** shows that school-based assessments collect information about school- and individual-related factors. In contrast, household-based surveys gather information related to progression and completion of education and other aspect of individuals in the household but naturally do not collect data about school-related factors.

### 7.1.2 What information do learning assessments collect?

**Figure 7.2** presents an overview of current availability of data/information for SDG 4 indicators from 22 existing learning assessments:

- 11 school-based learning assessments: EDI, EGRA/EGMA, ICCS, ICILS, LLECE, PASEC, PILNA, PIRLS, PISA, SACMEQ and TIMSS.
- 11 household-based learning assessments: East Asia-Pacific Early Childhood Development (EAP ECD) Scales, Education Health Center Initiative (EHCI), IDELA, ITU, MELQO, MICS, PAL network, PIAAC, PRIDI, STEP and Young Lives.

In total, these assessments account for 36 SDG 4 indicators.

- 5 indicators (4.1.1, 4.6.1, 4.4.2, 4.7.4, 4.7.5) require assessment by means of a module or test;
- 30 indicators could be sourced from information available in background questionnaires;
- 7 indicators don't have source information from the existing 22 assessments.

#### How can a country find examples of questions?

To complement this guide, a visualisation with the inventory of learning assessment survey questions from existing instruments helps to guide countries (and stakeholders) with examples on how to frame the question and what indicators the question is informing.

**Figure 7.2 List of SDG 4 indicators that can be sourced from each assessment**

| Type of assessment | Assessment | |
|---|---|---|
| School-based | TIMSS | 4.1.1 4.2.1 4.2.2 4.a.1 4.c.1 4.1.3 4.1.6 4.1.7 4.2.3 4.2.4 4.S.2 4.a.2 4.c.2 4.c.3 4.c.4 4.c.7 |
| | PASEC | 4.1.1 4.2.2 4.a.1 4.c.1 4.1.6 4.1.7 4.2.4 4.5.2 4.S.4 4.a.2 4.c.2 4.c.3 4.c.4 4.c.5 4.c.7 |
| | PIRLS | 4.1.1 4.2.1 4.2.2 4.a.1 4.c.1 4.1.3 4.1.7 4.2.3 4.2.4 4.5.2 4.a.2 4.c.2 4.c.3 4.c.4 4.c.7 |
| | SACMEQ | 4.1.1 4.2.2 4.a.1 4.c.1 4.1.3 4.1.6 4.2.4 4.5.2 4.7.2 4.a.2 4.c.2 4.c.3 4.c.4 4.c.6 4.c.7 |
| | PISA | 4.1.1 4.2.2 4.a.1 4.c.1 4.1.6 4.1.7 4.2.4 4.5.4 4.7.5 4.a.2 4.c.2 4.c.3 4.c.4 4.c.7 |
| | TERCE | 4.1.1 4.2.2 4.a.1 4.c.1 4.1.6 4.2.4 4.5.2 4.a.2 4.c.2 4.c.3 4.c.4 4.c.7 |
| | ICCS | 4.7.1 4.a.1 4.c.1 4.1.3 4.1.7 4.2.S 4.7.4 4.a.2 4.c.2 4.c.3 4.c.4 |
| | EGMA/EGRA | 4.1.1 4.2.2 4.a.1 4.c.1 4.2.4 4.S.2 4.c.2 4.c.3 4.c.4 4.c.7 |
| | ICILS | 4.4.1 4.a.1 4.c.1 4.1.3 4.4.2 4.c.2 4.c.3 4.c.4 4.c.7 |
| | EDI | 4.2.1 4.2.2 4.2.4 |
| Household-based | Young lives | 4.2.1 4.2.2 4.3.1 4.a.1 4.1.3 4.1.4 4.1.5 4.1.6 4.1.7 4.2.3 4.2.4 4.2.5 4.3.2 4.3.3 4.4.3 4.S.2 4.S.4 4.a.2 4.c.7 |
| | MICS | 4.2.1 4.2.2 4.3.1 4.4.1 4.1.3 4.1.4 4.1.5 4.1.6 4.1.7 4.2.3 4.2.4 4.2.5 4.3.2 4.3.3 4.4.3 4.5.2 4.6.2 |
| | PAL Network | 4.1.1 4.2.2 4.3.1 4.a.1 4.c.1 4.1.3 4.1.4 4.1.5 4.1.6 4.1.7 4.2.4 4.5.2 4.5.4 4.6.2 4.c.2 |
| | STEP | 4.2.2 4.3.1 4.4.1 4.6.1 4.1.3 4.1.4 4.1.5 4.1.6 4.2.4 4.3.2 4.3.3 4.4.3 4.S.2 4.6.2 4.6.3 |
| | PIAAC | 4.3.1 4.4.1 4.6.1 4.1.4 4.1.5 4.3.2 4.3.3 4.4.3 4.6.2 |
| | EAP ECO Scales | 4.2.1 4.2.2 4.2.3 4.2.4 4.2.5 4.5.4 |
| | EHCI | 4.2.1 4.2.2 4.2.3 4.2.4 |
| | IDELA | 4.2.1 4.2.2 4.2.3 4.2.4 |
| | MELQD | 4.2.1 4.2.2 4.2.3 4.2.4 |
| | ITU | 4.4.1 4.4.2 |

*Source: UNESCO Institute for Statistics (UIS), http://gaml.uis.unesco.org/dashboard*

The visualisation is a comprehensive table showing the indicator concept, name, number, the type of assessment, the assessments that measure it, and in which questionnaire the question can be found. An information "i" icon is available for each assessment, and the question is presented with a hover-over functionality over the icon.

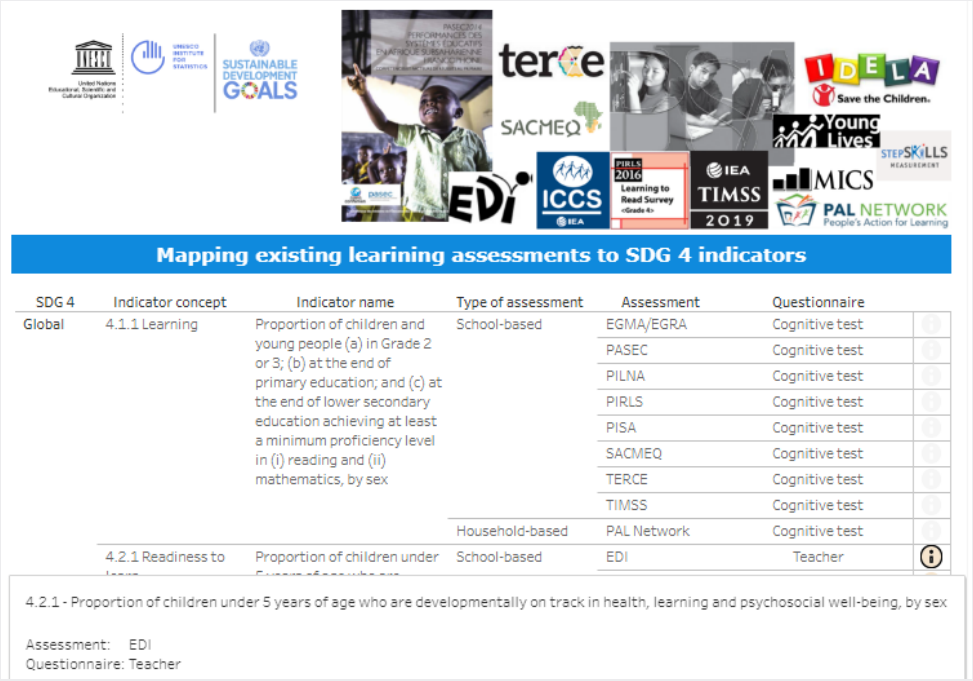
This exercise would help both country-level and international actors to gain a vital new set of tools that could support them in tracking and achieving inclusive and equitable quality education and the promotion of lifelong learning opportunities for all.

## 7.1.3 Can learning assessments serve to measure equity?

The results show that about 60% of inequality is within countries, and "only" 40% is between countries. The "between" component would likely grow if the international income inequality increased, but the most important source is within countries. Assuming a goal of reducing worldwide inequality, it is a priority for countries to work on the factors that could help to reduce the "within country" component of inequality.

The comparison by sex, wealth and location are relevant, but there are other factors that reinforce or do not help to attenuate the situation. For example, socioeconomic status implies a less-beneficial impact of home life for children of low status than for children of higher socioeconomic status. As families self-select, some types of schools of lesser quality and less preparedness to support learning reinforce the circle. This is the greatest source of inequality in countries (Crouch and Rolleston, 2017; Crouch and Gustafsson, 2018).

**Figure 7.3 Mapping existing learning assessments to SDG 4 indicators**




*Source: http://gaml.uis.unesco.org/dashboard*

It is vital to understand all sources of inequality. Evidence shows that disadvantaged pupils in socioeconomic terms apparently attend lower-performing schools and schools which are less effective but are also affected by greater uncertainty with regard to school performance. Improving equity means focusing more attention/resources on the disadvantaged, including focusing on pure inequality or lack of "standards", such as mastery of a (realistic) curriculum or quality and appropriateness of teaching and books (rather than using sex, location or income as proxies). Understanding how factors interact and how this issue affects the most disadvantaged in terms of cognitive skills are likely to be the most productive approach to improving equity.

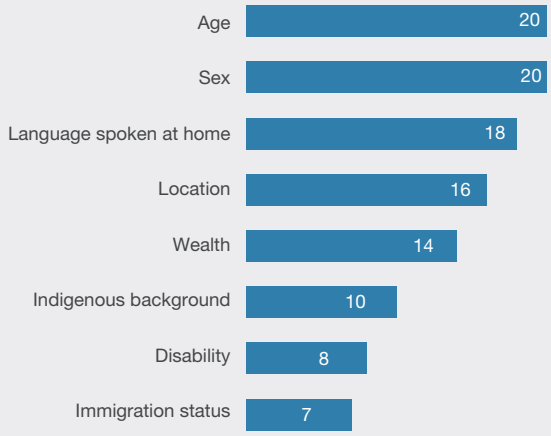
Learning assessments could provide a unique tool to help understand those aspects. Both in-school and household survey-based assessments host a large amount of information across all SDG 4 targets *(see **Figure 7.4**)*:

- Disaggregation by age, sex, home language, location, socioeconomic status, indigenous background, immigrant status and disability are extensively found in the existing learning assessments but are not covered with the same intensity.
- Age and sex information are available in all 20 assessments examined.
- Ethnic background, immigrant status and disability information are found to be the least available among current learning assessments.

Other aspects related to inequality, such as teacher training and school environment, could also be collected by in-school learning assessments. Information on these factors are key to understanding how schools reinforce or reduce learning gaps between advantaged and disadvantaged students.

A remaining issue is that in many cases the questions are not necessarily comparable *(see **Box 7.2**)* and

**Figure 7.4 Availability of disaggregated data out of a total of 20 assessments**




*Source: UNESCO Institute for Statistics (UIS), http://uis-azr-prod-wordpress-eus1.azurewebsites.net/gaml/capacity-development/*

this could be explored in several of the background questionnaires. Challenges related to comparability occur within assessments over time and between questionnaires. Background questionnaires are often collected from different actors (school head, teacher, student, family) in the same assessment and results may be very different. Questions can also vary over time within the same instrument, as well as varying across instruments. Even between instruments used at the country level, questions relating to broad topics such as school services and resources may vary (e.g. ASER and Uwezo, both citizen-led assessments under the PAL Network ask different questions on the available school services).

## 7.2 HOW TO IMPLEMENT A LEARNING ASSESSMENT IN MY COUNTRY?

Different learning assessments have different purposes and characteristics. Countries aiming to introduce learning assessments should be aware of these differences and should select the assessment that best fits their national education goals, needs and resources. In all likelihood, more than one type of assessment may be needed, especially if one counts both highly-formal assessments and less-formal assessments used in the classroom or by ministry providers of quality assurance services to schools.

Learning assessments may measure different subject areas at different levels and grades of the education cycle. They may vary in the frequency of administration and on the costs of implementing the assessment. National assessments are usually better-fitted for measuring the national curriculum, whereas regional and cross-national assessments

### Box 7.2 How do learning assessments define location?

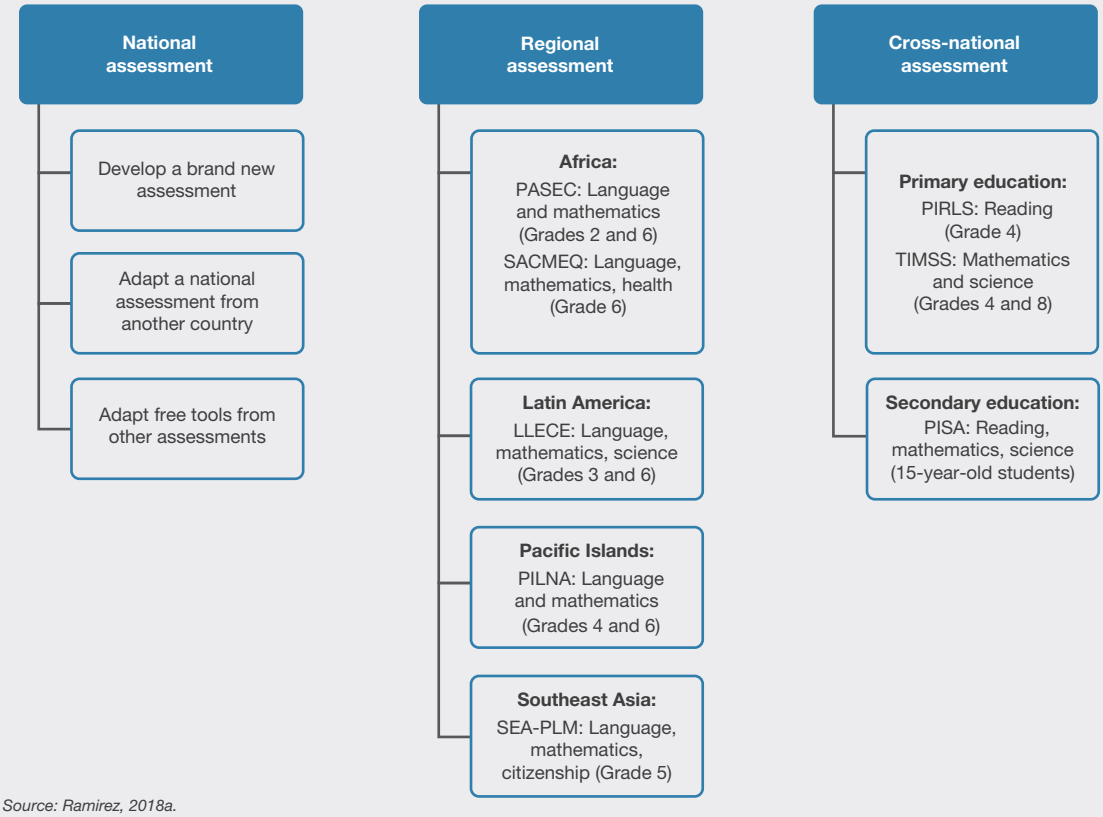
Slight distinctions appear when examining school locations. Large-scale assessments provide no information about the location of students' homes. They only provide basic information regarding the location of schools.

Another issue relates to the definition of rural and urban areas. In some assessments, the distinction between locations of schools is based on the number of people living in the area, while in other assessments the definition is more subjective.

In PISA, information concerning the definition of location is provided in the questionnaire to avoid potential misunderstanding of rural/urban areas. On the contrary, the distinction between locations of schools is more complicated for PASEC and SACMEQ assessments. This difficulty may be explained by the geographical structure of sub-Saharan African countries. For instance, the questionnaire asks school directors if the school is located in "a city", "a suburb of a large city", "a big town" or "a small town". It may be complicated to differentiate between "a big town" and "a suburb of a large city". The same observation can be made for SACMEQ, where it is not clear how to differentiate between an "isolated area" and a "rural area".

*Source: UNESCO Institute for Statistics (UIS).*

**Figure 7.5 Options to consider when deciding what type of learning assessment to implement**




*Source: Ramirez, 2018a.*

allow for making international comparisons. Regional assessments are available for countries that usually share common geographic, cultural, linguistic or historical backgrounds; they are usually available in the subject areas and grades that are considered a priority for those regions. Cross-national assessments have a more global presence and their results have been considered for reporting in the beginning stages of SDG 4 *(see Chapter 2).* As methodologies are developed, all assessments will report to SDG 4 in a harmonised way so that they are comparable.

### 7.2.1 Options for implementing a national assessment

As shown in **Figure 7.5**, countries that decide to conduct school-based national assessments can follow different strategies. One strategy is to develop a brand new assessment. This strategy is the most common and has the advantage of greater ownership by stakeholders. A new assessment usually ensures better alignment with the national curriculum, which is important when reporting if students are reaching the curriculum objectives.

Another strategy is to adapt a national assessment already being used in another country. For instance, Mozambique put in place its national assessment by adapting an assessment programme from Brazil *(Provinha Brasil)* to measure reading in the first cycle of primary education. This South-South collaboration saved Mozambique time and resources by not having to "reinvent the wheel".

A third strategy that countries should consider is to adapt learning assessments that are freely available

online, i.e. that are part of the public domain. Instruments and procedures from assessment programmes, such as EGRA and EGMA, ASER and Uwezo, can be used to monitor learning in mathematics/numeracy and language/literacy, and are available in different languages. For example, Gambia has been administering its own local adaptation of EGRA/EGMA to nationally-representative samples of students. Pakistan has been administering its refined version of the ASER household test annually to nationally-representative samples of children and youth. This approach has the advantage of offering free and ready-made tools for measuring learning. The disadvantage is less ownership by stakeholders and less flexibility to address national curriculum considerations. One possible disadvantage of simply borrowing downloaded tools is the loss of the technical assistance and quality assurance that accompanies the development of such tools. However, this kind of assistance can be obtained cheaply or for free through bilateral or multilateral development agencies.

Different assessments can complement each other. For example, a country may administer a national assessment in Grade 3 and a cross-national assessment in Grade 6. However, given scare resources, countries may have to opt for one or the other. To economise, the country may use a much more informal assessment, still capable of producing useful information but not of the accuracy of the Grade 6 assessment, at Grade 3 level. (However, minimum levels of reliability and validity need to be assured.)

Figure 7.5 shows the options a country may consider when deciding what type of learning assessment to implement. The first decision concerns whether to conduct a national, regional or cross-national assessment. There are pros and cons for each of these assessment types *(see **Table 7.1**),* and countries should weigh them according to their own local context.

### 7.2.2 Key stages in implementing a learning assessment

**Figure 7.6** presents the key stages of the assessment cycle that should be taken into account during implementation. Most of these stages apply to all assessment types (national, regional or cross-national), with different emphasis and somewhat different activities. Implementing the entire assessment cycle may take around three years, although this varies considerably from country to country.

To inform Indicator 4.1.1, countries will need to provide evidence that each one of these stages was implemented, meeting technical criteria.

### 7.2.3 What are the alternative institutional arrangements for a learning assessment unit?

Different institutional arrangements are possible to implement a national, regional or cross-national assessment. Common arrangements include:

- *Unit within the ministry or department of education*. This is probably the most typical arrangement. Having the learning assessment unit within the ministry or department has the advantage of facilitating coordination among curriculum, pedagogy and assessment teams. This facilitates alignment among these components and uses of the assessment results. It has the disadvantage of being more vulnerable to political interference and corruption from high levels or from colleagues and peers including teachers (e.g. not publishing or altering poor results).

- *Semi-autonomous public institution*. Several countries have national institutes of statistics, research centres or quality assurance agencies leading national, regional or cross-national assessments. These institutions have their own budget and are accountable to the minister of education or congress. They have the advantage

**Table 7.1 Pros (+) and cons (-) of national assessments vis-à-vis cross-national and regional assessments**

| National assessments | Cross-national and regional assessments |
|---|---|
| Politics | |
| (-) More likely to be affected by country politics. Results may not be published or they may not be trusted. | (+) Independent of country politics. Results are more likely to be trusted |
| Local stakeholders | |
| (+) Allows for involving local stakeholders in the assessment and therefore more likely to ensure their support. | (-) Local stakeholders are less involved in the assessment and therefore may be less likely to support it. |
| Curriculum | |
| (+) Usually more aligned with the national curriculum and its learning objectives. | (-) Usually less aligned with the national curriculum and its learning objectives. |
| Capacity | |
| (-) The local team may not have access to appropriate training to implement the assessment. | (+) The local team can benefit from high quality, hands-on training in each step of the assessment. Very valuable to build local capacity. |
| Costs | |
| (+) May be cheaper than a cross-national assessment. Countries need to cost for test development, data analysis and reporting. | (-) May be more expensive than a national assessment. Countries need to cost for participation fees and travels, and assessment implementation. |

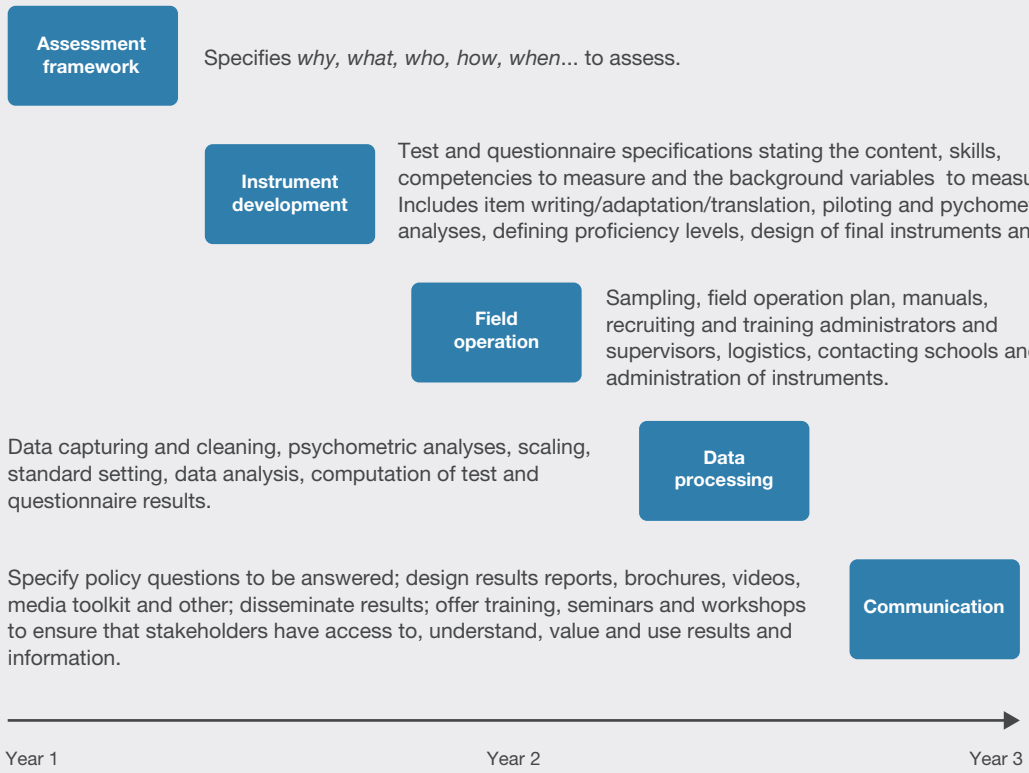
*Source: UIS, 2017d.*

## Box 7.3 Main challenges when conducting a large-scale assessment

- *Failure to secure political support and stable funding.* As a consequence, the assessment stability is at risk. Involving stakeholders and transparency are essential to minimise this risk.
- *Need to secure sufficient staff.* It is important to agree on the number of staff and the amount of time they put into the assessment, and to plan the assessment accordingly.
- *Need to develop local capacity.* The best way to do so is by providing hands-on training while implementing an assessment. Efforts should be made to retain the trained staff.
- *Poor sampling.* Sampling and fieldwork should be planned in detail and well in advance.
- *Lack of standardised procedures.* Manuals, training and quality control procedures are important tools to ensure standardisation.
- *Assessment results are not comparable.* Major issue when the aim is to report changes in learning across years. In design, make sure it has the technical features needed for comparing results.
- *Assessment results are not published.* Planning for a communication strategy where poor results are used as a baseline to promote improvements.
- *Lack of an assessment culture.* Produce simple reports, flyers and websites that address a few key research questions, and offer workshops to explain results and the assessment in general.

*Source: Ramirez, 2018a.*

**Figure 7.6 Stages and activities typically needed to implement a learning assessment**




*Source: Ramirez, 2018a.*

of being more independent of political or collegial influence. The risk is a lack of coordination and misalignment of the assessment with other components and policies of the education system (e.g. the assessment team not communicating with the curriculum team to ensure that the tests measure the curriculum objectives).

- *Examination board or unit*. Countries that have examinations for certification (e.g. secondary school diploma) or selection purposes (e.g. university entrance examinations) may benefit from having the same institution in charge of national, regional or cross-national assessments. The advantage is the benefit of the institutional capacity and expertise of the examination team. A disadvantage is the possibility of overwhelming an institution that already has a clear mandate.
- *Outsourcing to a university, NGO or equivalent*. Some countries outsource the implementation of their national, regional or cross-national assessments. They do so by forming strategic alliances of five or ten years, or by signing contracts with one or more institution to be in charge of the whole or a part (e.g. field operation) of the assessment. This arrangement is more common for regional and cross-national assessments. In national assessments, it is used more often during the initial introduction of the assessment or during the first years of its implementation (e.g. a university is in charge of a national pilot assessment). A limitation for low- and middle-income countries is the lack of local institutions with the technical capacity to implement the assessment.

Regardless of the arrangement, the institution leading the assessment should be accountable to a clearly recognisable body (e.g. the minister of education, congress or a national education commission) that is itself accountable.[91]

### 7.2.4 How much does a learning assessment cost?

A national learning assessment may cost between US$200,000 and US$1,000,000, depending on several factors, such as target population (in-school versus out-of-school children/youth), number of students tested, administration mode (e.g. group versus individual administration), local costs of services (e.g. printing) and personnel (e.g. test administrators). It is important to estimate total costs and secure sufficient and stable funding (e.g. from government and donors). Perhaps the most important cost factor is whether the sampling frame of the assessment allows specific inference about performance of sub-national jurisdictions (states, provinces) given that a larger sample size is needed for valid results at the sub-national level. In other words, sometimes the results are valid for the national aggregate and allow statistical and policy inferences at that level, but the sampling does not allow a conclusion for some disaggregation, either at the sub-national level or according to another classification.

The cost of regional and cross-national assessments is likely to be around USD$1,500,000. Again, this may vary greatly depending on the assessment programme, implementation plan and local costs (UIS, 2018a).

When comparing the financial costs of the large-scale assessment to the cost of running an education system for an "average" country, we realise that a large-scale assessment is an investment. According to the UIS database, the average cost for low- and middle-income countries to run their pre-primary to secondary education system is about US$5.8 billion per year. If we assume (as studies have shown) that education systems have at least 10% inefficiency, the average cost of inefficiencies in a country would be around US$580 million per year. If 5% of this inefficiency, in a conservative scenario, is addressed by having and properly using learning assessment data, then the benefit is about US$30 million per year in an average country.

With an estimated annual cost of US$250,000 for two assessments every four years, the benefit/cost ratio would be 30/0.25=US$120 million per year. This exercise would produce stunning results at the global level. For every 100 countries that invest, the benefits are clear (100 * US$120 million per year). In other terms, the approximately US$1 million invested every four years for an assessment (or US$250,000 per year) amounts to just one-tenth of 1% of the running costs for the entire education system (UIS, 2018a). Any modern organization spends at least that much on quality control systems, relative to its revenue.

## 7.3 HOW TO SHARE AND DISSEMINATE LEARNING ASSESSMENT DATA?

To ensure that learning assessment data are used to the maximum extent possible, findings must be disseminated in an appropriate manner considering the intended audiences. In each country, the project team should create a dissemination plan that specifies key findings to be disseminated, identifies key audiences to be targeted and describes the dissemination or media approaches best suited for both the information and the audience.

In most countries, the major issue surrounding communications is the inability to inform teachers. Communicating results to teachers implies presenting the implications of the assessment in a highly-actionable manner. Unfortunately, most assessment units in ministries lack the required skills. Close collaboration with teacher trainers, principals and often committees of highly-senior and experienced teachers, together with curricular experts, is needed to design ways to inform and help teachers. However,

91 See UIS, 2018a.

this does not happen often, which is why it can take countries more than a decade before the assessment can positively impact learning outcomes. This cycle can be shortened substantially if conscious decisions and budgets are devoted to doing so.

### 7.3.1 Objectives of learning assessments

**Learning assessments will play a key role in informing SDG 4**. Most importantly, they can play a key role in supporting learning outcomes and therefore SDG Target 4.1. For this to happen, it is necessary to ensure the effective use of learning assessments.

**Measuring learning is not an end in itself**. We measure learning to improve learning. However, it is not enough to administer a test to the students and to report assessment results. It is necessary to ensure the effective use of assessment information. For stakeholders to make effective use of assessments, they need first to have *access* to the information, to *value* and *understand* it. They also need some *contextual and institutional conditions* that allow them to use assessment information to support learning.

**There is concern about the under-utilisation of assessment information**. Sometimes countries make a tremendous effort to administer a national assessment. Different subject areas are measured, questionnaires are administered, but then results are not published or they are only disseminated in an internal report within the ministry or department of education. In other cases, results are more broadly disseminated but still may fail to reach key audiences, such as teachers. Another challenge is that reports are written in a technical language that does not resonate with educators.

**Contextual factors may also hinder the effective uses of assessment results**. Educators usually complain about the lack of time to examine assessment results. Teachers may be overwhelmed by other teaching and administrative tasks that take priority over reading assessment reports. Supervisors may not have the capacity to follow up on the assessment results of all schools under their jurisdictions.

**In general, there has been more concern about implementing national assessments than in using them effectively**. This is understandable considering the enormous technical, institutional, financial and political challenges that arise with the introduction of a new assessment. However, there is now a push towards ensuring effective use of the assessment information. The focus of attention is changing from the production of assessment data to their effective utilisation. The next section provides some guidelines on this subject.

### 7.3.2 Target audiences

The chosen methods and media for dissemination will depend on the target audiences and their priorities and levels of interest. Typical audiences include:

- *Non-technical ministry of education officials, local donors and NGOs working in the country's education sector.* These stakeholders will be most interested in key findings and summary statistics to inform the design, evaluation, continuation or termination of education programmes and reforms, to decide about funding allocation and to distribute incentives to schools or other stakeholders.

- *School principal and supervisors.* They are interested in monitor learning at the school level, to set learning targets and to provide tailored support for teachers to reach those targets. Other activities such as implementing workshops for teachers to analyse, understand and use assessment results would benefit from effective communication and dissemination.

- *Teachers.* This is one of the most important stakeholder groups. Assessment results would help to complement their own information and adapt pedagogy to the learning needs of students that reach different proficiency levels, to

## Box 7.4 Lessons from the Kenyan Tusome programme

The Kenyan national literacy programme, Tusome Early Grade Reading Activity, uses student learning data in three ways to support improved literacy instruction. First, the national teacher professional development programmes use the results of EGRA literacy assessments to inform teachers whether the country's Grade 1 to 3 students have reached the benchmarks for learning at each grade level. These benchmarks, set by the ministry of education, are reinforced at each training, helping the teacher to calibrate their expectations for improved learning and noting progress of their county and even their classroom towards those benchmarks.

Second, curriculum support officers who serve as coaches in the system visit literacy classrooms to support teachers implementing the Tusome programme. Their visits are focused on instructional quality and they provide feedback to teachers on the specific instructional practices needed to improve quality using a tablet-based system. At the end of each visit, the coaches randomly select three students and undertake a simple literacy screening measure. The results of these assessments are shared with each teacher during every visit, which provides generalised instructional feedback from a particular lesson within the context of the achievement of the students in that classroom.

Third, the data from each of these individual classroom observations are uploaded to the cloud and a national dashboard of instructional quality and learning outcomes reflects those findings. Given the scope of the programme, this means that results from more than 60,000 student literacy assessments are available every month, and the educational leaders in Kenya can compare results over time and across geographical areas. The purpose of all of these learning assessment opportunities is to help teachers be aware of the learning levels in their classrooms and to improve the quality of their instruction in general, and specifically to those learners who need more help.

*Source: Piper et al., 2018.*

understand contextual aspects and comparisons that might be of help. It could help teachers to ensure special support for students that do not reach minimum proficiency levels.

- *Parents, family.* This critical stakeholder group could support student learning at home if they understand the different factors involved. They could also support schools and exert more pressure and demand greater accountability of schools. They could make more informed decisions on school selection.

- *Students*. They can use assessment results as both intrinsic and extrinsic motivation to improve their own results, as well as those of their peers. In addition, students can work closely with teachers and parents to keep school administration and governments accountable and push for greater funding and support for their schools.

- *International or multilateral donor organizations.* These stakeholders will be interested in cross-national comparisons of indicator measurements as well as the full technical report and dataset. They allocate funding.

- *Academic researchers and technical units within the ministry of education.* A small audience will require a full technical report. More specifically, curricular units, lesson-planning support units and textbook re-design units need to communicate with assessment units and set up programmes that help teachers. According to Crouch and Rollerstone (2017), the main source of inequality is sheer variance itself, due to a lack of clear, specific, useful standards that teachers can implement. Researchers and technical units could play a pivotal role in improving standards in teaching to increase the impact of learning assessment.

- *General public and media.* Average citizens and the media, both local and international, should be able to easily access key findings. They can aide in keeping schools and governments accountable, while helping to improve learning environments and safety.

### 7.3.3 Dissemination formats

Various audiences will require different levels of detail, which can drive decisions about the dissemination format. Various means of communicating are listed below, and **Table 7.2** maps the target audiences to different dissemination formats.

- *Infographics, briefs and other non-technical materials* that can be accessed online or printed for dissemination events. They present key findings that are clear and concise, with minimal text *(see **Figure 7.7**).*

- *Policy briefs* that connect specific survey findings with related policy implications. For example, if data from a household survey show disparities in educational attainment between boys and girls, a policy brief can show the relevant data points (such as pre-primary attendance, primary and secondary completion rates and participation in technical-vocational training programmes) to highlight the significance of the disparity at various levels. In consultation with subject experts (e.g. gender) and key stakeholders (e.g. ministry of education official for primary grades), authors of the brief can then connect the data to suggested policy changes that could lessen disparities. See an example from IEA's Compass (see ***Figure 7.8***).

- *Online dashboards.* Each of the above can be made publicly-available online, provided all necessary permissions are granted by government and funders, in the form of static PDF files or an interactive dashboard. SDG indicator data can be added to repositories of education data – online dashboards – that are interactive in nature and allow comparisons with other country data, such

**Table 7.2 Stakeholder dissemination tools**

| | Stakeholders | | | | | | |
|---|---|---|---|---|---|---|---|
| **Type of communication** | Policy-makers | Principals and supervisors | Teachers | Parents and families | Academia and NGOs | Media | General public |
| Event/ presentation | ✓ | ✓ | | | | ✓ | ✓ |
| Briefs and infographics | ✓ | ✓ | | ✓ | | ✓ | ✓ |
| Online data dashboard | ✓ | ✓ | | | | ✓ | ✓ |
| Technical report | | | | | ✓ | | |
| National/sub-national school report | ✓ | ✓ | ✓ | | | | ✓ |
| Assessment framework | | ✓ | ✓ | | | | |
| Dataset | | | ✓ | | ✓ | | |
| Social media/ website | ✓ | ✓ | ✓ | ✓ | | | ✓ |
| Videos | | ✓ | ✓ | ✓ | | ✓ | ✓ |

*Source: UNESCO Institute for Statistics (UIS).*

**Figure 7.7 Example of an infographic**


*Source: UNESCO Institute for Statistics (UIS).*

**Figure 7.8 IEA's Compass policy brief**


*Source: IEA Compass Briefs in Education, No. 2.*

as UNESCO's eAtlas for Education 2030 *(see* ***Figure 7.9****)*. Such dashboards enable even non-technical users to visualise data, creating charts, graphs and maps showing data from a single year or across years.

- An *oral presentation* covering key findings, preferably accompanied by a visual component, such as a PowerPoint presentation.

- *Media.* In many places, radio and television remain good outlets for highlighting key survey findings. In addition, once dissemination materials are created, the general public and news media can be alerted to them via social media. For example, key findings from India's 2016 ASER are described in a seven-minute video[92] or the UIS explains in a three-minute video why data are needed to help get all children in school and learning by 2030.

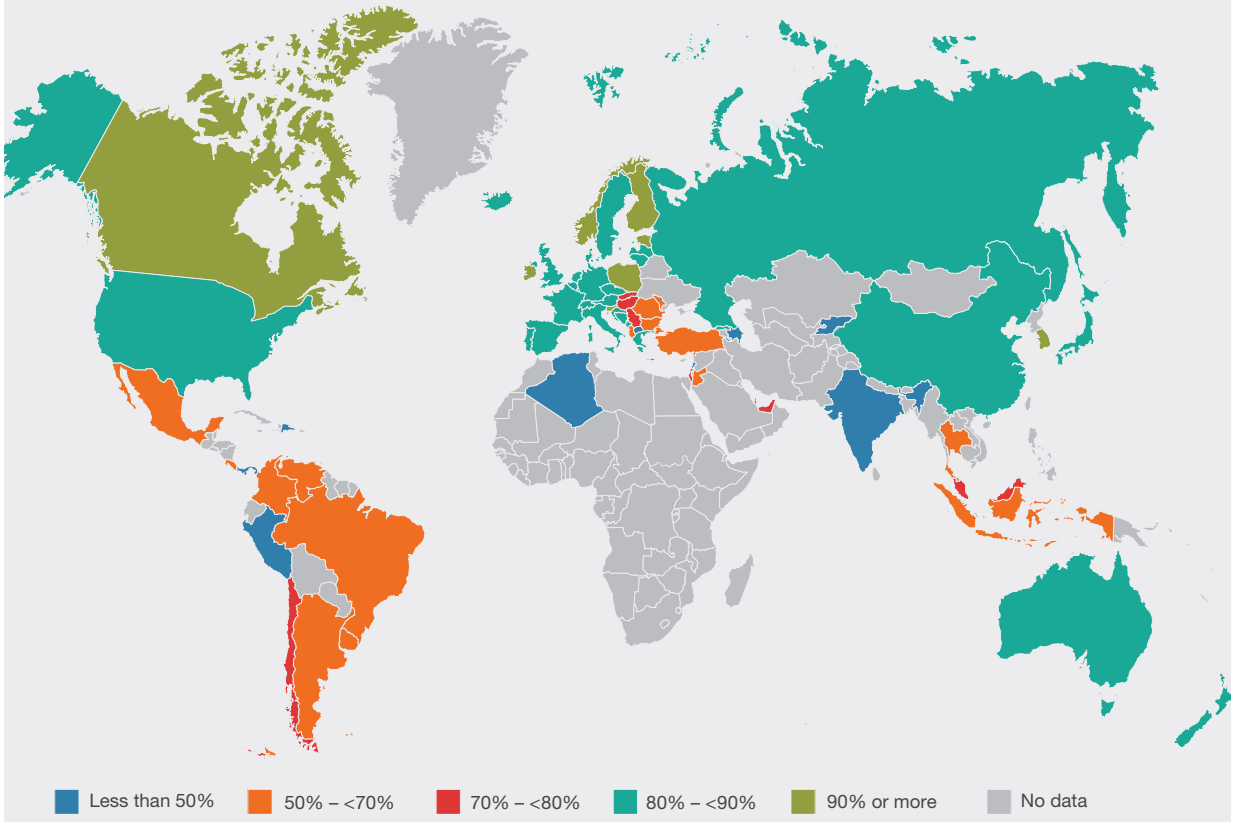
- *Social media.* Twitter could be effective in communicating results, as well as Facebook, and other platforms can be used to help further disseminate findings and materials. They could be very effective as discussed in Box 7.5. Instead of or in addition to printed reports, a series of WhatsApp messages could be sent with infographics or animations showing the results, example questions/items of the tests and links to videos with pedagogical resources. Clear and simple messages, with specific guidelines for action, could be sent to parents, teachers, principals, supervisors and policymakers once a week, over a period of several weeks or months.

- *Reports on technical findings or factsheets* that include, at a minimum, sub-sections describing the purpose of the assessment, the methodology applied (sampling, instrument development, fieldwork and data analysis), the findings (which should be clearly tied to the SDG 4 indicators), any limitations of the approach and a discussion of implications for achieving SDG 4 in light of the findings. This type of analysis is available in the UIS fact sheet on children not learning (2017g).

- *Datasets.* The full dataset must be shared with the relevant unit within the ministry of education (as well as with the funding entity if distinct from the government). In addition, the cleaned and de-identified dataset and codebook can be made into public use files that will be useful to researchers,

92 ASER Centre, 2017

**Figure 7.9 Example of an indicator map in the UIS eAtlas for Education 2030**




*Source: UNESCO Institute for Statistics (UIS).*

both nationally and internationally, who wish to conduct secondary analyses. This can be made available through a secure electronic transfer process that requires verification of the person or group requesting the dataset, as well as their intentions for its use. Subsequent findings from secondary analyses should also be disseminated using similar platforms.

### 7.3.4 Recommendations

Carrying out a learning assessment from conception to completion is a large and complex undertaking. Those embarking on this endeavour will face political, financial, technical and logistical challenges. Being aware of potential issues can help to avoid challenges entirely or mitigate their impact on the task of producing high-quality learning assessment data.

Countries can use learning assessment data to inform SDG 4 and take advantage of the rich data collected via in-school and population-based assessments. These data can greatly assist in countries' efforts to monitor progress towards both international and national goals. Yet, there are some pending tasks to address regarding learning assessment data.

It is true that there are challenges related to comparability within assessments and between them but they are still a unique source of information. Combining the information from assessments with other sources and building an integrated information management dashboard would make it possible to design strategies for "mass learning" that might include minimum (and quite specific) standards of schools/learning, teachers, management and pedagogy in order to grant the minimum for every

child and youth. This should be accompanied by specific forms of accountability and support to meet those standards (as opposed to generic support such as more pro-poor spending).

No one should be left behind. The reasons for poor performance are key and at the heart of the SDG global and thematic framework. Both household surveys and in-school assessment could help in understanding the links between home, school and disadvantage and to feed action to solve them.

More information and information on topics yet to be explored in assessments and surveys could be collected. Currently, there is no school-based assessment targeting the upper secondary education level (youth aged 15 to 17 years). Household questionnaires do not collect information on illiterate populations or on the use of skills. There are very limited data on global citizenship education in schools, as only the ICCS national context survey covers this topic.

## Box 7.5 Learning assessments and social media in Paraguay

Paraguay administered a national learning assessment to all students and schools in 2015. School results were published and disseminated to all departments, provinces and schools in 2018. What follows is a real conversation between a learning assessment specialist and a rural teacher from Paraguay. This conversation took place in the context of the evaluation of the communication strategy of SNEPE, the national learning assessment of Paraguay.

Assessment Specialist: *Did you see this school report with the SNEPE results for your school?*

Teacher: *Nooo... First time I see it ... maybe the school principal got it... but we, the teachers, we didn't...*

AS: *And the principal did not share or distribute it with the teachers?*

T: *The school does not have a photocopy machine, and we don't have computers or printers... Moreover, there is no internet connection here, so it is very hard for us to have access or to share the SNEPE reports.*

AS: *Mmm... I see... we would need to send printed copies to every teacher then...*

T: *Why don't you send us a WhatsApp?*

AS: *A WhatApp???!!!*

T: *Yes... we don't like to read printed reports... We prefer to read from our cell phones. All the teachers have one, and we have WhatsApp groups. If you send us a WhatsApp, everybody would be informed of the school results! We could even share with the parents; they use WhatsApp too!*

AS: *Really???!!!*

T: *Would it be possible to send WhatsApp messages in Guarani [indigenous language] for the parents?*

The conversation quoted above reflects the reality in many schools in Paraguay, especially the poorest and more isolated ones. Paradoxically, the lack of a conventional communication infrastructure pushed them to rely on new technologies in order to break their isolation. The main way of communication for these communities is social media. It is not printed reports or official memos. It is not the telephone, not even email. It is WhatsApp.

*Source: Ramirez, 2018b.*

# 8. Communications, uses and impact of large-scale assessments

Empirical research from around the world demonstrates the critical role of education in helping people lift themselves out of poverty, improve their quality of life, strengthen their health and that of their family, while increasing their employment opportunities and contributing to the economic development of their country. As a result, the international community has been striving to set educational goals and overcome challenges in reaching them, particularly in low- and middle-income countries.

As shown in previous chapters, it is extremely difficult to monitor learning outcomes globally because not all countries conduct national assessments or participate in regional and cross-national assessments. This poses a significant challenge in providing initial information for SDG 4 monitoring and reporting.

In addition, many low-income countries are not interested in participating in cross-national assessments, which they believe are too difficult for their children and therefore do not provide relevant information on the learning conditions in their countries. At the same time, the donor community does not have relevant information and quality data to inform their decisions on how to best support low-income countries to improve the learning outcomes of their children. It is therefore essential to provide the information needed by the international community to understand the value of advocating for and helping countries to develop and conduct national and cross-national assessments.

This chapter discusses the use of data from large-scale assessments. The first section describes the meta-analysis of existing literature on use, while Section 8.2 documents the potential uses for learning and Section 8.3 reviews the experience of the IEA.

## 8.1 THE IMPACT OF LARGE-SCALE ASSESSMENTS

A previous UIS study on the impact of large-scale assessments (UIS, 2017d) shows countries that have

### Box 8.1 About the synthesis

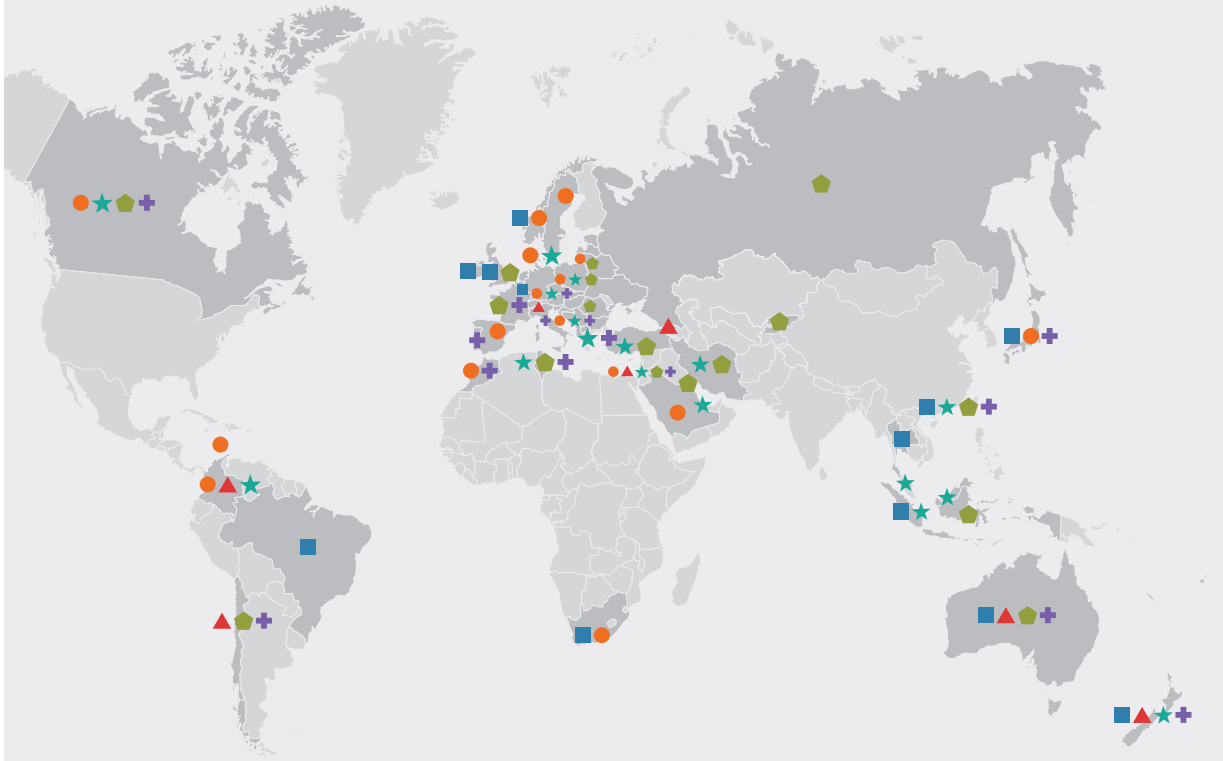
This synthesis is based on the UIS discussion paper entitled, "Review of the Use of Cross-national Assessment Data in Educational Practice and Policy". It reviews education policy and practice published since 2000 in order to present a relevant and timely synopsis of results without replicating key findings.

The synthesis addresses three key questions:

a. How do countries participating in cross-national (regional and international) assessments use their data for policy development?
b. What resources have countries invested based on the outcomes of a cross-national assessment?
c. What are the factors that prevent or hinder these countries from using the assessment information to improve education policies and outcomes?

*Source: UNESCO Institute for Statistics (UIS).*

**Figure 8.1 Effects of cross-national assessments on education policy**




*Source: UNESCO Institute for Statistics (UIS), 2017d.*

been benefiting from the results of cross-national assessments. From a policy perspective, the results of the review identify significant benefits arising from the use of cross-national assessment data. They include the use of data for comparative and benchmarking purposes; improving a country's overall education system through directive policy; enhancing access and equity; improving teaching and learning practice; curriculum reforms; and utilising strategies and indicators to monitor and evaluate education processes. **Figure 8.1** shows how countries can benefit from the use of cross-national assessment data, as in the cases of Canada, the United States and Australia.

### 8.1.1 How large-scale assessments guide investment

Large-scale assessment data have inspired new and creative forms of resource allocation in various countries. **Table 8.1** groups the examples of resource investment under three main umbrellas.

#### Teachers, training and professional development

Effective teaching depends on both the skills and motivation of teachers. Because both can be strengthened and developed, greater resource allocation for teachers has been a top policy priority as a result of international assessments.

**Table 8.1 Resources in which countries have invested based on the outcomes of cross-national assessments**

| Area of resource investment | Examples |
|---|---|
| Teachers, training and professional development | ■ New online in-service professional development programmes for teachers and leaders.<br>■ Teacher training workshops/integrating technology into classroom activities.<br>■ Incentives to participate in in-service teacher training programmes, encouraging high-performing students to join the teaching profession through incentives and increasing salaries.<br>■ Improving teachers' pedagogical skills and teaching literacy.<br>■ Incentives for teachers. |
| Education funding | ■ Increasing budget for education to provide primary and secondary education with additional financial resources to reduce class size, raise teacher salaries and develop infrastructure.<br>■ Several initiative investments to strengthen literacy development, including a generous Quality Education Fund.<br>■ Funding programmes to promote reading and literacy.<br>■ Donors helping to stimulate a policy response in terms of resource allocation in part through the administration of the assessment.<br>■ Interventions based on the findings, which are also used to influence policy dialogue and action. |
| Education materials and time resources | ■ An increase in classroom instruction time dedicated to mathematics leading to improved assessment scores.<br>■ Reductions in teacher shortages as a result of policy changes and efforts.<br>■ Hybrid assessment data being incorporated into a national assessment system to inform curriculum and instruction.<br>■ Hybrid assessment data to inform the development of materials and strategies for teaching and continuous assessment.<br>■ Influencing the national education programme, resulting in the allocation of significant funding to the building of classrooms, providing instructional materials, and addressing out-of-school children through non-formal education programmes. |

*Source: UNESCO Institute for Statistics (UIS), 2018c.*

## Education funding

Funding for education has been a concern and priority for countries, especially as a result of the growing international awareness stemming from assessments like PISA, which highlight the resources that are dedicated to education systems. The World Bank (2018) suggests that as countries increase their budgets for education, they should "shift spending patterns" so that teachers gain the necessary resources they require to improve student learning.

## Educational materials and time resources

Infrastructure, the availability of materials and use of time inside and outside the classroom all have substantial influence on learning outcomes of students. Increased allocation of resources does not suffice in improving learning: it must be combined or informed by better use of resources.

An increase in resources often affects learning outcomes to a small degree. What is more important

is how resources are allocated and efficiently utilised through focused and accountable policy measures that are better able to address and create whole-system improvements, even when countries are limited by finances.

### 8.1.2 Barriers to using large-scale assessment data in policymaking

Although large-scale assessments can provide valuable information for countries in terms of comparison, there are instances in which a country participates in the assessment yet disregards or fails to use the results in education policymaking *(see Figure 2.2 for geographic distribution of large-scale assessments).* Below we present the common barriers that prevent the use of assessments in education policy and provide examples.

#### Lack of or poor dissemination of information

- Little awareness of assessment results due to weak dissemination.
- Assessment teams do not share findings in sufficient or salient ways that improve education system operations.
- Only education officials and policymakers have access to assessment data, resulting in little public awareness and pressure.

#### Limitations in assessment programme and analyses

- Difficulty in comparing results from one assessment programme to another.
- Uncertainty of data being recognised at the national level.
- Limited capacity of technical experts to analyse large-scale assessment data.
- Assessments not responsive to pressing policy concerns of a country's education system.
- Results not used to specifically target or develop interventions at the classroom level.

#### Weak assessment bodies and fragmented government agencies

- Assessment mechanisms, especially concerning information dissemination, are inadequate or insufficiently organized.
- Fragmentation and reluctance among relevant government bodies in handling data.

#### Political factors

- Violent conflict and political unrest influencing the implementation of assessment.
- Lack of political will.
- Lack of efforts to improve reading instruction at the primary level despite indications from the data.
- No acceptance of assessment results or no agreement on how to implement changes.
- Discrepancies in findings resulting in a policy stalemate.
- Data manipulation and corruption leading to policy inaction or misdirection.

## 8.2 INFORMING POLICYMAKING

The role of large-scale international studies for informing education policy has mainly relied on two approaches. One is to collect data on a myriad of school and classroom factors and determine the relationships of these factors with learning outcomes. The results of these analyses are used to support various national policies. The second approach is for countries to compare their results with those of other countries. The factors considered relevant to student success are grouped into a number of policy themes, which can be broadly categorised as school resources, accountability, school governance, teaching practices and selective schooling.

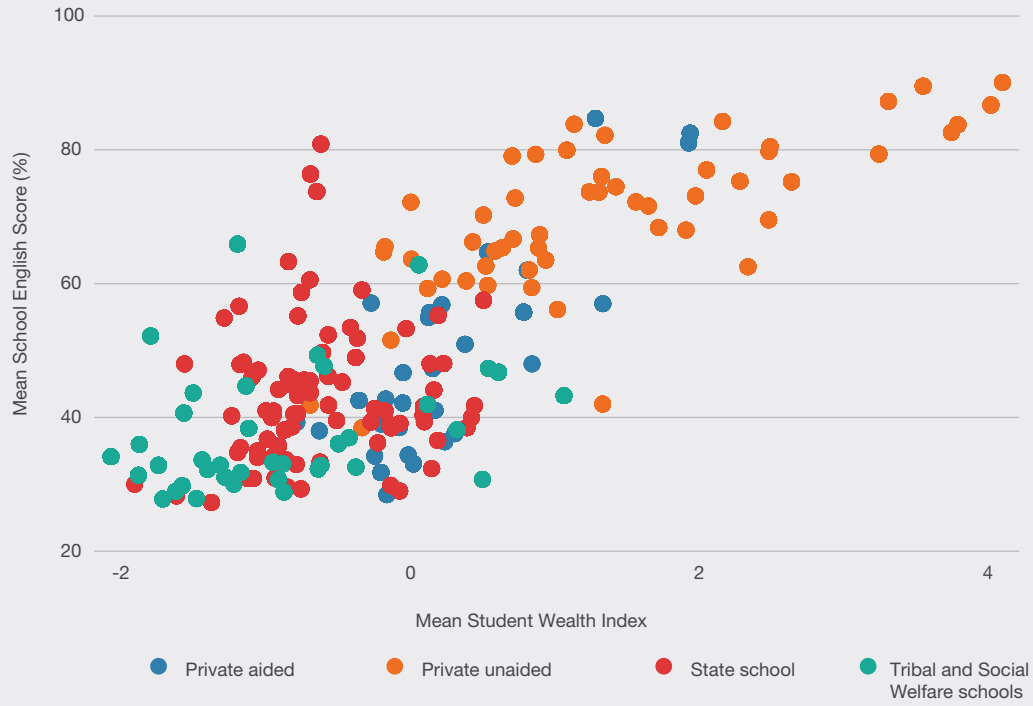
Both approaches are problematic and can fail to sufficiently address the *cumulative result* of countless factors that affect children's development, beginning at conception and continuing. Moreover, the measures of the key school factors that do affect student performance tend to be inter-correlated and

strongly correlated with the average socioeconomic status of the school. It is virtually impossible to isolate the "school effects" attributable to particular resources or processes with a cross-sectional study.

In an attempt to understand the relationship between pupils' home advantages and their school-level performance, Desa et al. (2008) examined learning outcomes in four main types of schools in India. While there is considerable variation in both the mean school English score and mean pupil assets score, there is a strong general pattern demonstrating that more advantaged pupils attend higher performing (and mainly private, unaided) schools. Overall, there is greater variation in academic performance for schools attended, on average, by more disadvantaged students. This finding is partly a function of the type of school attended (state schools have more variation in performance), but it is notable that within private schools there is more variance regarding performance than in schools attended by the most disadvantaged pupils. There is no discernible pattern among state schools, and there is no school type which, on average, has more advantaged pupils. In addition, a large proportion of state schools have lower performance than almost all private schools, due in part to their high concentration of disadvantaged students. In general, disadvantaged pupils attend lower-performing schools and schools which are less effective and also have greater variation regarding performance. **Figure 8.2** illustrates the general relationship between students' home advantage and their school-level performance.

The 2030 Agenda for Sustainable Development focuses on inequalities: gender, disability, immigrant status, language spoken at home and at school, and poverty, identified according to different databases.

**Figure 8.2 School-level performance by average pupil background, India**




*Source: Crouch and Rolleston, 2017.*

### 8.2.1 Simulating intervention effect

Monitoring data can inform policy questions about the performance of the school system and serve to monitor progress. This section considers three ways to frame questions about strategies and their execution.

For each strategy, the potential effect of a hypothetical strategy is also considered. In Figures 8.3 and 8.4 displaying the hypothetical effects, the red gradient line displays the "before intervention" status, which is set to the gradient, and the green line displays the hypothetical "after intervention" gradient.

### 8.2.2 Universal versus targeted strategies

**Figure 8.3** shows two different strategies, one universal and the other targeted at a certain level of performance. It summarises the learning bar in reading according to the socioeconomic status of the student. As expected, the level of learning increases with socioeconomic status as represented by the red line. The effect of simulation is shown by the green line.

The figure at the top shows a universal strategy which strives to improve the outcomes of all students in a jurisdiction. Reform could take place through different means: curriculum reforms, reducing class size, changing the age of entry into kindergarten or increasing the time spent on reading instruction are all universal strategies as they are targeted towards all students, irrespective of socioeconomic status. Consistently the bar in green simulates the effect of the universal policy.

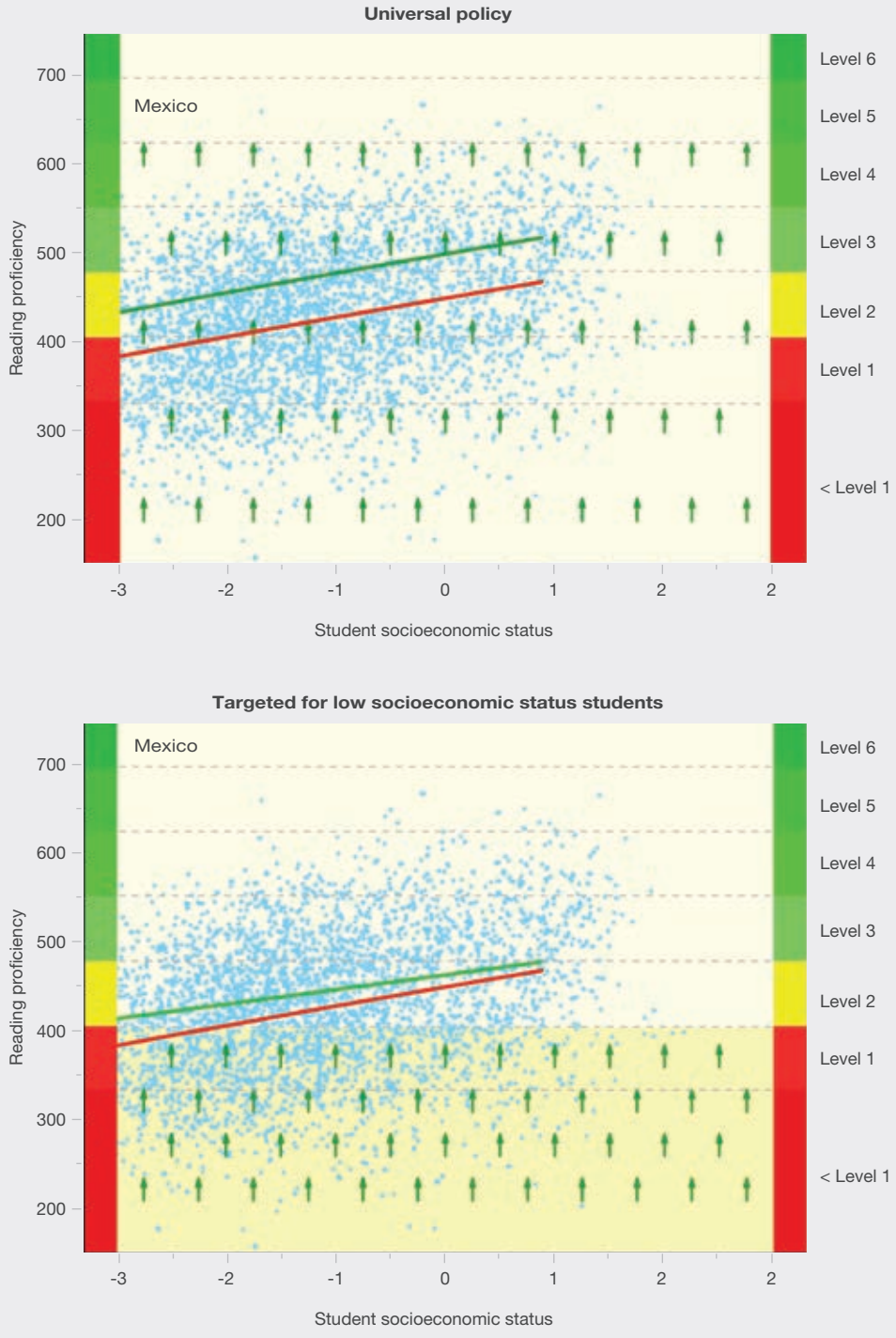
The figure on the bottom shows a strategy that is targeted towards students with low levels of performance based on an outcome in a performance-targeted intervention. Using sampling information from large-scale assessments, governments could target a school level by the type of school. A performance-targeted strategy can also be implemented at the school level. For example, a reading programme may be administered in a sample of schools that have low average performances. In school systems with a low vertical inclusion index, it is efficient to implement a whole-school strategy in a small number of schools. A vulnerability concentration plot can be used to estimate the number of children that would be reached with an intervention in a particular number of schools. The classification provides teachers with information regarding the *type and amount* of support required for each child.

### 8.2.3 Compensatory strategy

**Figure 8.4** shows the effects of a compensatory strategy that raises socioeconomic status. A compensatory strategy provides additional education resources to students from low socioeconomic backgrounds or students deemed "at risk" for other reasons. The term, "at risk" can refer to being at risk of not successfully achieving a particular development outcome or more generally at risk of poor development for a range of developmental outcomes.

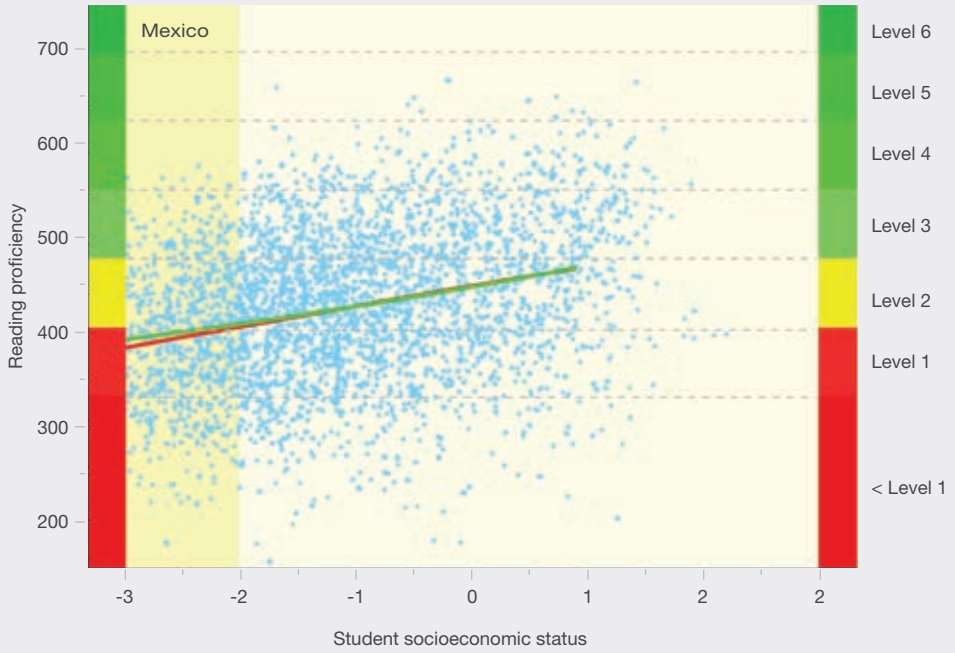
What is the difference between this intervention and the targeted one? The targeted sub-population can be the same as for an socioeconomic status-targeted intervention. A compensatory strategy intends to improve learning outcomes by improving the socioeconomic situation by using social and economic policy. This is a difference with respect to previous examples where the "compensation" comes from educational tools and policies as a way to improve outcomes. Providing free breakfast or lunch programmes or free textbooks for low socioeconomic status students are compensatory strategies. However, the effect is difficult to measure and it is indirect and dependent on many other factors. So far, no marked effect on improving children's outcomes has been proven.

**Figure 8.3 The effect of a universal and a targeted policy**




**Note:** The red bar represents the pre-intervention learning bar in reading according to socioeconomic status. The green bar represented the post-intervention learning bar.
*Source: Willms, 2018.*

**Figure 8.4 Simulating the effect of a compensatory policy**




**Note:** The red bar represents the pre-intervention learning bar in reading according to socioeconomic status. The green bar represents the post-intervention learning bar.
*Source: Willms, 2018.*

## 8.3 THE USES AND IMPACT OF IEA STUDIES[93]

The founders of the IEA viewed the world as a global laboratory, where different educational systems governed by national policies and practices produce different educational outcomes (Keeves, 2011). This diversity creates opportunities for comparative pedagogical research to test theoretical hypotheses and to investigate common problems across national educational systems. The evidence obtained from studying a wide range of educational systems reveals important relationships that would have otherwise remained unnoticed in a single system.

Since 1969, when the IEA embarked on the more extensive Six Subjects Study, it has produced comparative studies of academic and practical value. Over the IEA's 60 years of activity, its membership has changed from mainly scholars to institutions that represent a broad spectrum of educational stakeholders. This variety supports the IEA's rigorous scientific approach but also creates a demand for innovation in conducting and communicating research. Instead of stand-alone studies, today most IEA studies have a cyclical design, generating repeated measures, which help to understand educational systems and how they change over time. Despite these changes, the IEA's goal has remained constant: to understand and to improve education around the globe.

As well as collecting up-to-date information on the achievement levels of students at specified grades, subjects and cross-curricular areas, the IEA studies collect considerable background information on how educational systems provide educational opportunities to their students, as well as the factors that influence how students use these opportunities. The IEA uses this approach to improve education systems by informing practices, policy and research.

93 Written by the Executive Director and Senior Research and Liaison Adviser, IEA.

### 8.3.1 Benefits of taking part in an IEA study

Participating in an IEA study gives educational systems an opportunity to identify challenges, to see what interventions work and to share practices and learn from other participants.

IEA studies are designed to take the complexity of educational systems and the inputs of multiple stakeholders into account. As a result, the data gathered are relevant and useful for a variety of applications. The robust, carefully-designed instruments, rigorous procedures and quality control measures ensure high-quality, comparative standards and the reliability and validity of the data.

The IEA's mathematics, reading and science trend assessments are unique in the international study space because they are curriculum-based, examining what students are expected to learn (intended curriculum), what is actually taught in schools (implemented curriculum) and student outcomes (achieved curriculum). We believe that this is the most rigorous and fair approach to comparing educational systems. In addition, this approach provides practitioners and stakeholders with the range of information needed for evaluating and shaping their educational policies and practices. The IEA works with the national research coordinators of participating countries to ensure that what is tested is appropriate for their students. IEA studies also explore factors at the home level, school level and other areas related to learning. Results from IEA studies should be analysed and interpreted within this context.

IEA research data may provide deeper insight into topics such as equity in education, gender disparities, parental engagement and strategies, influence of student attitudes, and the other contextual factors linked to educational achievement.

In addition, engagement with IEA studies provides an opportunity for participants to build their own capabilities for educational assessments. Many countries use the experience and knowledge gained by participating in IEA studies to set up their own national assessments. In contrast, other countries such as the Czechia have decided against developing national tests. Instead, they implement international assessments as the only nationwide assessments of learning achievements and use those results to inform national educational agendas.

### 8.3.2 Sharing IEA's results and information

For each study, the IEA produces an international report that provides extensive high-quality information on students' achievement outcomes and their educational contexts and helps countries to assess their educational systems in a comparative context. Countries that have participated in previous cycles of the same assessment may also gain insights into their own national trends, as well as the international trends illustrated from the longitudinal data collection. Reports are available online, allowing readers to search and download or print particular topics and information of their interest.

The IEA's international databases allow for public access to the data collected and processed by each of its studies. All participating countries contribute by releasing their national data as part of these databases. The databases provide student achievement data, as well as background information about curricula and learning environments: students, home (in the case of PIRLS and TIMSS), teachers and schools. All data are anonymised so that scores cannot be linked to individual students or schools.

Alongside the main international report, our supporting publications help stakeholders to understand and work with these data. User guides describe the organization and content of each database. This complements the assessment framework's theoretical overview of the study and the documentation describing the rationale for the techniques used and the variables created in the process of data collection and compilation (technical report or methods and procedures). Each study's encyclopedia also gives

information from participating countries on the structure of their educational systems and school curricula. In order to work effectively with any of IEA's data, it is necessary to consult these publications to understand the characteristics of the study.

In addition to supporting researchers by providing data and documenting procedures, the IEA invests in methodological research with potential to provide new insights in the field of methodology and interpretations. To share its work and increase the impact of the research, the IEA invests in open source publishing of articles in the IEA-ETS Research Institute Journal, large-scale assessments in education and books as part of the Springer, *IEA Research for Education* series. It also offers workshops and training courses to support researchers working with large-scale assessment data. This commitment to making IEA's research publicly available has been an effective strategy for increasing the reach of findings.

IEA studies are familiar to many scholars and other education specialists within the subject areas of reading, mathematics, science, civics and citizenship, information and communication technology and teacher education. Its studies are also well-known by experts interested in methodology and statistical analysis of large-scale data on educational achievement worldwide. They are less well known by researchers in other fields, many practitioners and decisionmakers responsible for educational policy, particularly in the countries that are not represented among IEA membership. The IEA is addressing these issues through its commitment to publishing findings in open datasets and by actively sharing and promoting its results as a solid evidence base for researchers, educators and policymakers worldwide.

IEA data are recognised by UNESCO as invaluable for monitoring progress toward SDG 4, which encompasses a wide range of aspects related to education. Its longitudinal datasets include information about student achievement in core subjects (literacy, mathematics and science). This is in addition to contextual information about learning environments, access to education and the development of cross-curricular competencies in areas such as civics and citizenship and digital literacy *(see Section 3.2.1)*.

### Supporting educators

Once a study is completed, there are many resources that can be used by teachers and other stakeholders in classrooms and schools. Teachers, teacher educators and researchers may access some of the test items used in the assessments to understand what tasks students are expected to accomplish. They may look for items by content domains, such as algebra and geometry, or cognitive domains, such as knowing and reasoning. In addition, they can see what percentage of tested students across the participating countries answered each question correctly to gain insights into how and where students may be struggling. Background data almanac files contain weighted summary statistics for each participating country on each variable in the student, home, teacher and school context questionnaires, including the context questionnaire scales. This approach helps to identify the obstacles where students are not performing as well as expected and the potential interventions to address those challenges.

### Advantages of a curriculum-based approach

The IEA "curricular model" enables practitioners, educators and researchers to review, interpret and utilise results in their national context. This often starts with the production of a national report directed by, and directly used by, the educational community and decisionmakers within a single country. In some cases, countries team up in order to analyse matters of their common interest or to focus on differences and similarities among them (see *Northern Lights on TIMSS and PIRLS 2011* as an example).

### Linking research to policy and practice

Both international and national findings can lead to direct interventions and policy changes based on

study results, especially by teacher educators who shape teaching practice at a national level.

For the data to have a direct impact on pedagogy in the classroom, researchers and local teachers usually need to collaborate. One example is the Oxford University's PIRLS for Teachers project that aimed to use PIRLS data to provide teachers in England with guidance on improving their own teaching of reading in primary schools. In response to a request from the teachers for visual materials that they could display in a staff room, the project produced two posters about best teaching practices that were intended to prompt teachers to reflect on the methods they used to teach reading. The first poster encouraged oral after-reading activities – such as talking with peers and answering questions about what they have read – to foster interest in reading and motivate students who had limited exposure to books at home. The second poster summarised analyses revealing that boys were more motivated for reading and lower achieving pupils were more engaged with reading when these groups were exposed to a variety of reading resources (Hopfenbeck and Lenkeit, 2018).

While it is encouraging to see the influence of IEA study results on good practice at the classroom and school levels, findings must also reach and influence policymakers to achieve a more lasting impact. As an independent, non-political organization, the IEA does not make specific policy recommendations for individual education systems. The country-specific context and culture demand a lot of insider knowledge; from identifying the right questions to analysing, interpreting and understanding the results, while taking national influencing factors into account. The IEA supports countries by facilitating knowledge sharing and offering inspiration for potential evidence-based pedagogical and policy interventions. The *IEA Compass: Briefs in Education* series – formerly known as the *IEA Policy Briefs* – is made up of short, accessible articles published on its website and aimed at a general audience. The goal of the series is to connect study findings to recurrent and emerging questions in educational debates at the international and national levels to provide an evidence base for practitioners who are engaged in developing solutions for their own, national educational challenges.

Outcomes of IEA studies have influenced educational policy across its member countries. **Table 8.2** provides an overview of some of these changes. The changes can be grouped into four main areas: curricular changes, teacher's education, professional development and support, focusing on a specific group or a specific need of students, and material supports like textbooks, libraries, and other forms of physical mechanisms that support pedagogy.

In some cases, it is also possible to document where IEA study results led to the launch of a new agenda for an education system. For example, Germany has developed a dedicated digital agenda for education after ICILS 2013 results revealed a relatively low achievement level for students accompanied by a lack of computers in schools and adequately trained teachers (Fraillon et al., 2014).

Most changes attributed to IEA study results involve curricular amendments. This can be understood by the fact that the researchers engaged in IEA studies are often scholars active in teacher education. TIMSS and PIRLS have proved to be the most powerful agents of change, particularly because their design allows for the monitoring of trends over time. This is especially important in developing education systems that are striving to achieve universal enrolment where children from the most socioeconomically-disadvantaged communities are usually the last groups to be reached. These children need particular attention and lessons from TIMSS and PIRLS have helped countries to develop tailored pedagogical tools to engage them, as documented in the TIMSS 2015 and PIRLS 2016 Encyclopedias (Mullis et al., 2016a, 2017a and 2017c).

For example, Morocco has made significant improvements to both the equity and quality of its education system which are demonstrated in the country's achievement scores over time *(see*

**Table 8.2 Examples of the impact of IEA study results on national education systems**

| Country | Action | Study |
|---|---|---|
| Belgium | Support policies for low socioeconomic status and immigrant students. | ICCS |
| Botswana | Curriculum amendments; guidelines for classroom testing. | PIRLS and TIMSS |
| Canada (Ontario) | Curriculum changes; more time for mathematics and reading instructions. | PIRLS and TIMSS |
| Chinese Taipei | Results used as one of the primary resources in evaluating the efficacy of mathematics and science education and curriculum development.<br>Focus on assistance for low-performing and disadvantaged students in mathematics and science. | TIMSS |
| Czechia | Series of teacher manuals developed based on the most common misconceptions and errors of Czech students. | TIMSS |
| Egypt | Curriculum amendments; introduction of new teaching methods fostering interaction between students and teachers. | TIMSS |
| England | Teacher training programmes to stimulate positive attitudes towards reading. | PIRLS |
| Hong Kong, SAR of China | Teacher training programmes and other initiatives to stimulate children's reading. | PIRLS |
| Hungary | Extending reading teaching to Grade 6. | PIRLS |
| Indonesia | Focus on second language learners. | PIRLS |
| Jordan | Revision of the mathematics and science curricula; use of released items in the development of textbooks; development of related teacher guides and trainings. | TIMSS |
| Latvia | Lowering the school entry age from 7 to 6; new guidelines for teaching primary grades. | PIRLS |
| Lithuania | In-service training for primary grade teachers aimed at improving their teaching methods. | PIRLS and TIMSS |
| Malaysia | Measures to address students' lack of opportunity for application of knowledge and to develop higher order thinking skills (HOTS), including teacher trainings, textbook reviews, and increasing HOTS items in national assessments; new curriculum since 2011. | TIMSS |
| Oman | Improvements of curricula and revision of learning outcomes; teacher training focused on question development according to the cognitive domains and incorporating them in classroom instruction. | PIRLS<br>TIMSS |
| New Zealand | Focus on reading achievement of Maori and Pacifica children. | PIRLS |
| Norway | More focus on reading instruction, including an earlier start of reading instruction. | PIRLS |
| Romania | Curriculum amendments; new teacher guides as well as new science text books issues; more emphasis on reading informational texts. | TIMSS and PIRLS |
| Russian Federation | Alignment of achievement goals with the frameworks of the international large-scale assessments. | PIRLS, TIMSS |
| Singapore | Focus on policies supporting lower-performing, lower-socioeconomic status students. | PIRLS, TIMSS |
| South Africa | Support programmes for school and classroom libraries. | PIRLS |
| Spain | Reading promotion. | PIRLS |
| United Arab Emirates | Sharing the best practices of teaching. | TIMSS and PIRLS |

**Note:** Further details may be found in the references (Aggarwalla, 2004; Elley, 2002; Gilmore, 2005; Schwippert, 2003; Schwippert and Lenkeit, 2012).
*Source: IEA.*

**Table 8.3 Progress in Grade 4 students reaching the TIMSS and PIRLS low- and high-achievement benchmarks in Morocco**

| Morocco | Advanced international benchmark | | High international benchmark | | Intermediate international benchmark | | Low international benchmark | |
|---|---|---|---|---|---|---|---|---|
| TIMSS | 2015 | 2011 | 2015 | 2011 | 2015 | 2011 | 2015 | 2011 |
| % of students | 0 | 0 | 3 | 2 | 17 | 10 | 41 | 26 |
| PIRLS | 2016 | 2011 | 2016 | 2011 | 2016 | 2011 | 2016 | 2011 |
| % of students | 0 | 0 | 3 | 1 | 14 | 7 | 36 | 21 |

*Source: http://timssandpirls.bc.edu/timss2015/international-results/timss-2015/mathematics/performance-at-international-benchmarks/percentages-reaching-international-benchmarks-across-assessment-years/ and http://timssandpirls.bc.edu/pirls2016/international-results/pirls/performance-at-international-benchmarks/trends-at-the-international-benchmarks/*

***Table 8.3***). Almost all children have access to school (net enrolment in Grade 1 is above 97%), and Morocco now administers diagnostic tests at the beginning of the school year to facilitate student grouping so that specific learning support programmes can be designed and implemented for students with similar difficulties. PIRLS findings also influenced Morocco's Reading for Success project which encourages children to read for pleasure. This approach has proved to improve children's motivation to read and help them advance their reading proficiency *(see Table 8.2).*

## 8.3.3 The challenges of using IEA data

While IEA studies strive to be a comprehensive, representative sample of student achievement within an education system, interpreting and working with the results have their own challenges.

Studies are only administered to a sample of students that has been approved to be demographically representative of the target population. Sampling is based on the most important features of an education system (such as the languages of instructions, types of schools, geographical locations, etc.) and is linked to an analytical plan. As a consequence, the results of the sample can be generalised to represent a full education system, or any sub-population supported by a sufficient number of cases.

IEA studies are designed for three primary groups of stakeholders: education practitioners, policymakers and researchers. Some results, like the impact of early learning activities or gender differences in engaging with children, are also interesting and applicable for parents. Students may also be interested in results themselves, such as descriptions of achievement levels. Finding effective channels to communicate the information and value of IEA results with these diverse groups is challenging across contexts.

Once the information has reached its intended audience, turning those findings and insights into informed actions is not an easy task. Policy changes and other interventions at an education system level take time, both to implement and to gauge a measurable effect. In addition, identifying causal relationships between one system's policies and achievement scores is not straightforward. Directions of influence are not always obvious, and additional, confounding factors can be present that are not covered by a study's methodological framework but still have an influence on education outcomes.

For example, evidence from PIRLS and many other non-IEA studies indicates that children learn best in their mother tongue. Consequently, many education systems advise that instruction should be available in a student's home language. However, this can be very challenging to implement if there is a scarcity

of instructional and reading resources in those languages. National history can also be an important factor, particularly in post-colonial countries where there is a well-established legacy and perceived higher value placed on the language of the colonising nation instead of indigenous languages. A recent IEA policy brief explored how understanding the effect of past and present language policies is important when interpreting international achievement differences across countries (Howie and Chamberlain, 2017).

### 8.3.4 Conclusion

IEA studies provide a valuable basis for insights into the achievement and progress of students across the globe. Their results give researchers, policymakers and practitioners evidence to help make informed decisions about how education systems should develop. However, education systems are complex organisms that serve diverse communities, purposes and needs. Understanding how they operate demands communication, collaboration, persistence and time. By linking research, policy and practice, the IEA helps to build a better-educated world.

## 8.4 REGIONAL CAPACITY DEVELOPMENT INITIATIVES

In this section, we focus on three regional programmes that aim to improve learning exchange among practitioners.

### 8.4.1 Teaching and Learning Educators' Network for Transformation (TALENT)

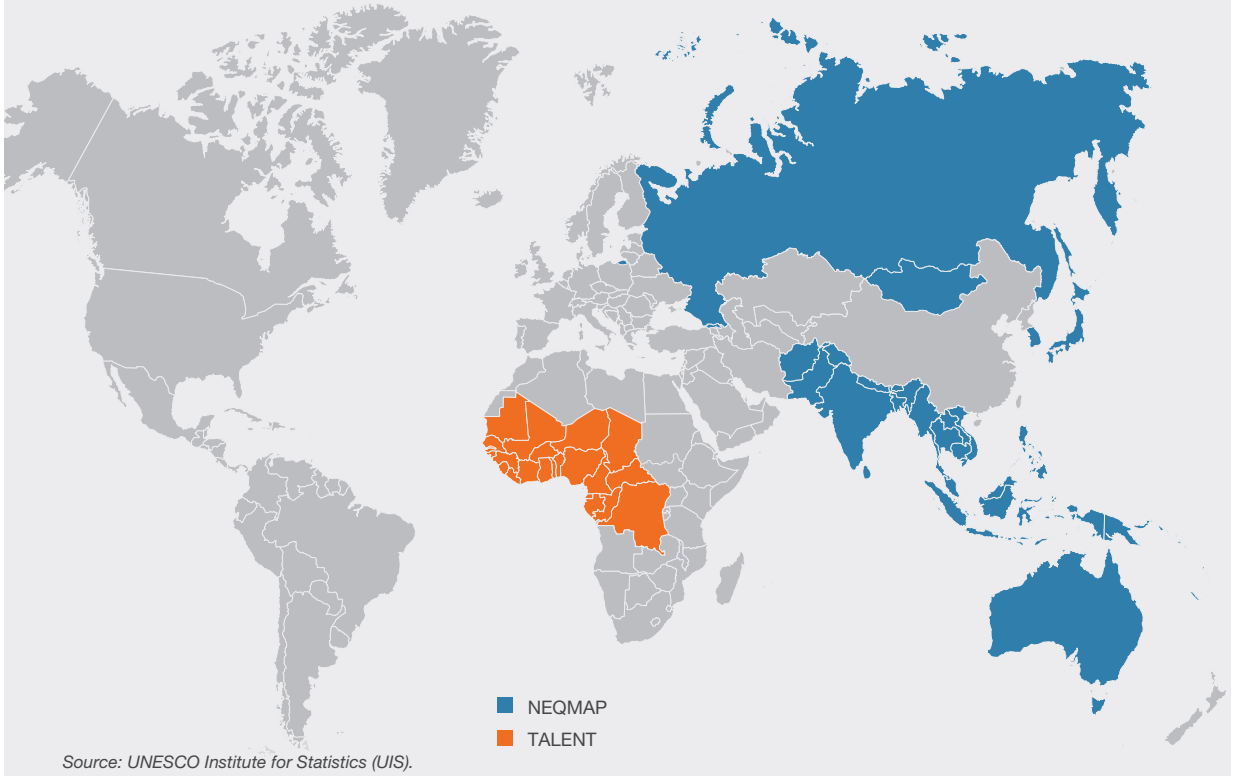
TALENT was established by the Regional Coordination Group on SDG 4-Education 2030 in June 2016 to serve as the platform for stakeholders engaged in regional programmes in sub-Saharan Africa to address learning crises. TALENT serves as a forum for exchanging experience, expertise and knowledge on initiatives to improve teaching and learning in the region, to promote research to inform policy changes and to develop capacity of stakeholders. UNESCO coordinates the network, while the UNESCO Office in Dakar acts as Secretariat, with the support of a steering group composed of members from the Association for the Development of Education in Africa - Network of African Learning Assessments (ADEA-NALA); Africa Network Campaign on Education for All (ANCEFA); CONFEMEN; *Réseau pour l'excellence de l'enseignement supérieur en Afrique de l'Ouest* (REESAO); and UNICEF. The steering group meets on a bimonthly basis to prepare work plans and to share, monitor and review expected outputs.

TALENT's focus area is teaching and learning, with particular attention given to curriculum alignment, assessment, teacher training and 21st century learning. The network's activities are founded on a theory of change to improve teaching and learning that includes three key steps: documentation of good practices and interventions, using national capacities, and networking of countries to enable South-South and North-South cooperation. Through a combination of interventions aimed at sharing experiences and best practices, producing and analysing evidence to inform policy and improving institutional capacities, the network's goal is to strengthen education systems in the region to ensure the acquisition of foundational, specialised and transversal skills by learners.

### 8.4.2 The Network on Education Quality Monitoring in the Asia-Pacific (NEQMAP)

NEQMAP, established in 2013 in Bangkok, Thailand, is a platform for the exchange of knowledge, experience and expertise on the monitoring of educational quality in countries and jurisdictions of the region. There are 25 member countries from the region and 7 associate members from different countries and organizations. The network focuses on student learning assessments, both national and large scale, as a tool to monitor education quality, while considering other enablers of classroom learning, including curriculum and pedagogy. UNESCO's Asia and Pacific Regional Bureau for Education (UNESCO Bangkok) serves as the secretariat. Through collaborative efforts, countries and jurisdictions share

**Figure 8.5 Geographical coverage of NEQMAP and TALENT**




*Source: UNESCO Institute for Statistics (UIS).*

lessons to improve the quality of learning in education systems with the eventual aim of influencing policy reforms.

NEQMAP activities concentrate on research, knowledge-sharing and capacity building among all stakeholders. The network works to enhance institutional capacities regarding learning assessments through a series of workshops on assessments and curriculum. The target audience is government officials responsible for designing and implementing national and large-scale learning assessments. In addition, the network provides technical support through workshops and expert reviews to institutions and/or countries that require specific help.

Recently, the network launched the NEQMAP Knowledge Portal, part of the National Education Systems and Policies in Asia-Pacific (NESPAP). The Knowledge Portal includes resources related to learning assessments, curriculum and pedagogy in the region, such as policy documents, research articles and reports, useful for government officials. NEQMAP also publishes a quarterly newsletter which covers topics related to learning assessments.

## 8.5 CIMA: IMPROVING EDUCATION DATA TO PROMOTE EVIDENCE-BASED POLICYMAKING IN LATIN AMERICA AND THE CARIBBEAN[94]

Education systems in Latin America and the Caribbean have made significant strides towards universal access and higher graduation rates in primary and secondary education (UNESCO, 2013). However, education quality is still very low and unequal, as revealed by the low performance of students in national, regional and international assessments and the wide learning gaps among students from different socioeconomic backgrounds.

94 Written by Elena Arias Ortiz, Florencia Jaureguiberry and Pablo Zoido, Inter-American Development Bank (IDB).

For example, 63% of 15-year-old students in the region do not achieve basic mathematics skills, almost three times as high as the 23% of low performers in OECD countries (Bos et al., 2016). This low performance in standardised learning assessments has been linked to poor economic performance (Hanushek and Woessmann, 2012).

In the last few years, many countries in the region have been actively pursuing innovative reforms and programmes aimed at improving student learning. However, strategic decisions, resource allocation and accountability are all fundamentally linked to the availability and adequate use of data (Burns and Köster, eds, 2016; Slavin, 2002). Unfortunately, the availability and reliability of data in many countries in Latin America and the Caribbean is uneven, and the difficulty of using the data to adequately inform education policy and practice can hamstring the efforts of policymakers to implement the reforms that education systems need.

To address these challenges the Education Division of the Inter-American Development Bank (IDB) launched CIMA *(Centro de Información para la Mejora de los Aprendizajes)* in 2016. CIMA is an education statistics portal that features comparable indicators on the education systems of most Latin American and Caribbean countries. CIMA's objective is to improve the collection, dissemination and use of education statistics in general, including learning achievement data. To achieve this goal, CIMA strengthens data systems and institutional capacity of education systems in the region. Collaborating with the governments of these countries, CIMA supports them technically and financially to: i) strengthen their systems of evaluation, data collection and analysis; ii) implement high-quality national learning assessments and participate in regional and international student learning assessments; and iii) evaluate the impact of any significant education reform.

### 8.5.1 CIMA's four pillars of action

The first pillar is an IDB-hosted portal of education statistics (iadb.org/cima) that presents more than 40 homogenised and comparable indicators describing the state of the 26 education systems in Latin America and the Caribbean. The website is available in Spanish, English and Portuguese in a user-friendly format, and graphs and tables can be easily downloaded in a standardised format. Comparable indicators for all countries with available data are organized in six categories: efficiency, coverage, physical resources, financial resources, context and learning. The CIMA website also features indicators by country. The indicators are calculated using three main sources of information: harmonised household surveys, administrative data (via countries and the UIS), and national, regional and international assessments.

The second pillar consists of a series of short publications, called *CIMA Briefs,* that highlight key trends shaping the quality and equity of education and learning based on data from the CIMA website. This series is organized around topics such as *Latin America in PISA*, describing the main highlights of the results of PISA*; CIMA Indicators Briefs*, analysing trends and current status of the key CIMA indicators; *CIMA Research Briefs*, drawing attention to selected data-driven analysis from different IDB education projects; and *CIMA Country Profile Briefs*, presenting country-specific data analysis.

The third pillar is to establish a CIMA network of government institutions that seek to improve the collection and use of education data for policy dialogue, design and implementation. The IDB works with these organizations to generate, validate and update key education indicators through a series of events and meetings aimed at facilitating peer learning, the exchange of policy experiences and closer cooperation across the region. CIMA supported the creation a working group dedicated to the study of composite education quality indicators in the region, in 2017 and 2018, alongside several

education evaluation agencies. CIMA has co-sponsored events in Quito, Santiago and Lima to share knowledge and experience among countries, experts and civil society organizations.

Finally, the fourth pillar is a series of capacity-building activities, driven by country-specific needs and priorities related to data-gathering and analysis. CIMA has hosted national workshops for government officials, on issues such as data harmonisation, while also supporting the participation of several Latin American and Caribbean countries in regional and international assessments, such as PISA for Development, in order to improve the quality of, and equity in, education.

CIMA statistics are used both within IDB initiatives and documents, as well as in work done outside the purview of the Bank. Data from CIMA were used and cited in loan documents in Uruguay, Ecuador, Panama, Honduras, the new Sector Framework Document and in the recently launched Development In the Americas (DIA) 2017 programme entitled "Learning Better". IDB education specialists use CIMA data regularly in their work, supporting client countries, conducting presentations and facilitating dialogues among governments and stakeholders. Outside of the Bank, CIMA has had a positive reception among journalists, researchers and policymakers. In a non-scientific survey distributed in August 2016 to selected users, 70% of respondents found CIMA's content relevant or very relevant for their work, and found it easy or very easy to interpret the data as presented. Among others, CIMA has been cited in regional media and other publications, such as the Ministry of Education, Colombia; *Diario El País*, España; *Diario ABC*, Paraguay; *Red Latinoamericana por la Educación* (REDUCA); *Blog Certeza*, Perú; *Efecto Cocuyo*, Venezuela; *Red TTU*, Colombia; and UN Economic Commission for Latin America and the Caribbean (CEPAL).

The harmonised information that CIMA gathers contributes to monitor at least five of the ten targets of SDG 4 in the region. CIMA contains information on early childhood development and pre-primary education, tertiary education indicators, school physical resources and quality of education through the analysis of national, regional and international assessments of students learning outcomes. Additionally, CIMA disaggregates all indicators, when possible, by sex, socioeconomic status, geographic location, school administration and financing source. Thus, CIMA is also a tool to monitor education systems' equality, an effort that is consistent with the 2030 Agenda premise of leaving no one behind.

In addition to CIMA, the IDB's Education Division has launched two other regional projects that directly aim to improve the use of data and evidence for decisionmaking in education: SUMMA *(Laboratorio de Investigación e Innovación en Educación para América Latina y el Caribe)* and New Leaders in Education. While CIMA focuses on gathering data and making it more readily available, SUMMA (www.summaedu.org) is a research and innovation lab for effective education policies created in 2016 in collaboration with *Fundación Chile* and the support of the Education Ministries of Brazil, Chile, Colombia, Ecuador, Mexico, Peru and Uruguay. For this purpose, SUMMA works in the following areas: i) generating knowledge and evidence through cutting-edge research on key matters of education policy; ii) boosting innovation in education through the promotion of policies that are innovative and have proven effective; and iii) stimulating the collaboration and exchange of knowledge between policymakers, academics, innovators and educators.

The second related initiative is New Leaders in Education, a series of online courses aimed at training policymakers and education stakeholders in the identification and use of evidence to inform education policy. Along with CIMA, these initiatives contribute to improving the capacity of education decisionmakers and key players to implement and mobilise proven education policies and programmes and, in alignment with SDG 4, thus contribute to improving the quality and equality of education systems in Latin America and the Caribbean.

# References


Aggarwalla, N.K. (2004). "Evaluation Report: Quality assessment of primary and middle education in mathematics and science (TIMSS)". Report to the United Nations Development Programme and the United Nations Office of project Services. New York: United Nations Development Programme (UNDP).

Ala-Mutka, K. (2011). "Mapping Digital Competence: Towards a Conceptual Understanding". Joint Research Centre (JRC) Technical Notes. Brussels: JRC, European Commission.

Altinok, N. (2017). "Mind the Gap: Proposal for a Standardised Measure for SDG 4–Education 2030 Agenda". UIS Information Paper No. 46. Montreal: UNESCO Institute for Statistics (UIS).

American Educational Research Association (AERA), American Psychological Association (APA) and National Council on Measurement in Education (NCME) (2014). "Standards for educational and psychological testing". Washington: APA.

Anderson K. and A. Raikes (2017). "Key Questions on the Domains of Measurement for SDG 4.2.1". Recommendations from GAML Task Force 4. Montreal: UNESCO Institute for Statistics (UIS).

Antoninis, M. and S. Montoya (2018). "A Global Framework to Measure Digital Literacy [Blog Post]". https://sdg.uis.unesco.org/2018/03/19/a-global-framework-to-measure-digital-literacy/

ASER Centre (2017). "ASER 2016: National Findings". https://www.youtube.com/watch?v=kUQxJjqa-o4

Black, M.M., S.P. Walker, L.C. Fernald, C.T. Andersen, A.M. DiGirolamo, C. Lu and A.E. Devercelli (2016). "Early Childhood Development Coming of Age: Science through the Life Course". *The Lancet*, Vol. 389(*10064*), pp. 77-90.

Bolly M. (2018). "Developing Evaluation Capacity in Africa: The Example of Action Research on Measuring Literacy Programme Participants' Learning Outcomes". Hamburg: UNESCO Institute for Lifelong Learning (UIL).

Bolly, M. and N. Jonas (2015). *Action Research: Measuring Literacy Programme Participants' Learning Outcomes. Results of the First Phase (2011–2014)*. Hamburg: UNESCO Institute for Lifelong Learning (UIL).

Bos, M.S., A. Elías, E. Vegas and P. Zoido (2016). "Latin America and the Caribbean in PISA 2015: How Many Students are Low Performers?" Washington: Inter-American Development Bank (IDB).

Burns,T. and F. Köster (eds.) (2016). *Governing Education in a Complex World, Educational Research and Innovation*. Paris: OECD.

Carr-Hill, R. (2017). "Improving Population and Poverty Estimates with Citizen Surveys: Evidence from East Africa". *World Development,* Vol. 93, pp. 249-259.

Carretero, S., R. Vuorikari and Y. Punie (2017). "DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use". Joint Research Centre (JRC) Report. Brussels: JRC, European Commission.

Chenery, H. and T.N. Srinivasan, eds. (1988). *Handbook of Development Economics,* Vol. 1 (1st edition). Oxford: Elsevier.

Cizek, G. and M. Bunch (2007). *Standard Setting.* Thousand Oaks, CA: Sage Publications.

Cole, M. and X.E. Cagigas (2010). "Cognition". In M. H. Bornstein (Ed.), *Handbook of Cultural Developmental Science*, pp. 127-142. New York: Psychology Press.

Courtois, L. (2013). “La recherche-action considérée comme un potentiel vecteur de changement des pratiques professionnelles du secteur associative: Le cas de la recherché visant la promotion de comportements citoyens“. Rome: AREF International Congress.

Crouch, L. and M. Gustafsson (2018). *Worldwide Inequality and Poverty in Cognitive Results: Cross-Sectional Evidence and Time-Based Trends*. Oxford: RISE.

Crouch, L. and C. Rolleston (2017). “Raising the Floor on Learning Levels: Equitable Improvement Starts with the Tail”. Paper presented at the Research on Improving Systems of Education conference, RISE Insight, London, U.K. https://www.riseprogramme.org/publications/raising-floor-learning-levels-equitable-improvement-starts-tail

Desai, S., S. Duby, R. Vanneman and R. Banerji (2008). “Private Schooling in India: A New Educational Landscape”. *India Policy Forum*, Vol. 5, Issue 1, pp. 1-58.

Dubeck, M.M. and A. Gove (2015). “The Early Grade Reading Assessment (EGRA): Its Theoretical Foundation, Purpose and Limitations”. *International Journal of Educational Development*, Vol. 40, pp. 315-322.

Educational Testing Services (ETS) (2014). “A Guide to understand the Literacy Assessment of the STEP Skills Measurement Survey”. Washington: World Bank.

Elley, W.B. (2002). “Evaluating the Impact of TIMSS-R (1999) in Low and Middle-Income Countries: An Independent Report on the Value of World Bank support for an International Survey of Achievement in Mathematics and Science”. Unpublished paper.

European Commission (2018a). “DigComp into Action – Get Inspired, Make It Happen.” https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/digcomp-action-get-inspired-make-it-happen-user-guide-european-digital-competence-framework

European Commission (2018b). “The Digital Competence Framework”. https://ec.europa.eu/jrc/en/digcomp

European Commission (2016). “A New Comprehensive Digital Skills Indicator”. https://ec.europa.eu/digital-single-market/en/news/new-comprehensive-digital-skills-indicator (Accessed July 2018).

Ferrari, A. (2013). “DigComp: A Framework for Developing and Understanding Digital Competence in Europe.” Joint Research Centre (JRC) Report. Brussels: JRC, European Commission.

Foy, P. (Ed.) (2018). *PIRLS 2016 International Database and User Guide*. Boston: TIMSS and PIRLS International Study Center.

Fraillon, J., W. Schulz and J. Ainley (2013). *International Computer and Information Literacy Study Assessment Framework*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Fraillon, J., J. Ainley, W. Schulz, T. Friedman and E. Gebhardt (2014). *Preparing for Life in a Digital Age. The IEA International Computer and Information Literacy Study International Report*. Cham: Springer.

Gaëlle P., M.L. Sanchez Puerta, A. Valerio and T. Rajadel (2014*). STEP Skills Measurement Surveys: Innovative Tools for Assessing Skills*. Washington: World Bank.

Gal, I. (2018). “Developing a monitoring scheme for adult numeracy as part of SDG indicator 4.6.1: Issues and options for discussion”. Discussion Paper for the UNESCO Expert Meeting on Adult Literacy and Numeracy Assessment Frameworks, 17-18 May 2018, Paris.

Gilmore A. (2005). *The Impact of PIRLS (2011) and TIMSS (2003) in Low- and Middle-Income Countries: An Evaluation of the Value of World Bank Support for International Surveys of Reading Literacy and Mathematics and Science*. Amsterdam: IEA.

Global Partnership for Education (GPE) (2018). *Results Report 2018*. Washington: GPE.

Global Partnership for Education (GPE) (2017a). “GPE 2020 Theory of Change”. Washington: GPE.

Global Partnership for Education (GPE) (2017b). *GPE 2015/2016 Results Report*. Washington: GPE.

Global Partnership for Education (GPE) (2017c). “Methodology sheet for GPE results Indicator 1.” Washington: GPE.

Global Partnership for Education (GPE) (2017d). “Methodology sheet for GPE results Indicator 15.” Washington: GPE.

Global Partnership for Education (GPE) (2016). *Results Report 2015/16.* Washington: GPE.

Goodnow, J. J. (2010). “Culture”. In M. H. Bornstein (Ed.), *Handbook of Cultural Developmental Science*, pp. 3-19. New York: Psychology Press.

Goodnow, J. J. (1990). “Using Sociology to extend Psychological Accounts of Cognitive Development”. *Human Development*, Vol. 33, pp. 81-107.

Gove, A., C. Chabbott, A. Dick, J. DeStefano, S. King, J. Mejia and B. Piper (2015). “Early Learning Assessments: A Retrospective”. Background Paper for Education for All Global Monitoring Report 2015. Paris: UNESCO.

Grotlüschen, A., D. Mallows, S. Reder and J. Sabatini (2016). “Adults with Low Proficiency in Literacy or Numeracy”. OECD Education Working Papers, No. 131, Paris: OECD.

Gustafsson, M. (2018). “The Costs and Benefits of Different Approaches to the SDG Indicator on the Proficiency of School Students”. UIS Information Paper No. 53. Montreal: UNESCO Institute for Statistics (UIS).

Hanushek, E. A. and L. Woessmann (2012). “Schooling, Educational Achievement and the Latin American Growth Puzzle”. *Journal of Development Economics*, Vol. 99, Issue 2, pp. 497-512.

Hopfenbeck, T. N. and J. Lenkeit (2018). “PIRLS for Teachers: Making PIRLS Results more useful for Practitioners”. Policy Brief No. 17. Amsterdam: IEA.

Howie S. and M. Chamberlain (2017). “Reading Performance in Post-Colonial Context and the Effect of Instruction in a Second Language.” IEA Policy Brief No. 14. Amsterdam: IEA.

Hungi, N., D. Makuwa, K. Ross, M. Saito, S. Dolata, D. van Cappelle, L. Paviot and J. Vellien (2010). “SACMEQ III Project Results: Pupil Achievement Levels in Reading and Mathematics”. Gaborone: SACMEQ.

IEA (2018). “The Data Makes the Difference: How Chinese Taipei Used TIMSS Data to Reform Mathematics Education”. IEA Compass: Briefs in Education, No. 2. Amsterdam: IEA.

Indian Market Research Bureau (IMRB) (2014). “National Sample Survey of Estimation of Out-of-School Children in the Age Group 6-13 in India”. Social and Rural Research Institute, IMRB and Educational Consultants India Limited (EdCil,) Delhi, September 2014.

Janus, M. and C. Reid-Westoby (2016). “Monitoring the Development of all Children: The Early Development Instrument”. *Early Childhood Matters 2016*, pp. 40-45.

Jeantheau, J-P. (2015). “La dictée dans les enquêtes sur la ‘littéracie’ des adultes : pratiques, résultats, exemples d’analyses, perspectives“. *Revue de sociolinguistique en ligne*, Vol. 26.

Jeong, J., A. Bhatia and G. Fink (2018). “Associations between Birth Registration and Early Child Growth and Development: Evidence from 31 Low- and Middle-Income Countries”. *BMC Public Health*, Vol. 18, Issue 1, pp. 673–681.

Joint Research Centre, European Commission (2018). “Self-Reflection Tool for Digitally Capable Schools”. https://ec.europa.eu/jrc/en/digcomporg/selfie-tool (Accessed July 2018).

Kariger, P., E.A. Frongillo, P. Engle, P.M. Rebello Britto, S.M. Sywulka and P. Menon (2012). "Indicators of Family Care for Development for Use in Multicountry Surveys". *Journal of Health Population and Nutrition*, Vol. 30, Issue 4, pp. 472–486.

Keeves, J.P. (2011). "IEA – From the Beginning in 1958 to 1990". In: Papanastasiou, C., Plomp, T., Papanastasiou, (Eds.) *IEA 1958-2008: 50 years of experiences and memories. Volume 1*. Amsterdam: IEA.

Kolen, M.J. and R.L. Brennan (2014). *Test Equating, Scaling and Linking: Methods and Practices*. Berlin: Springer Science and Business Media.

Lancaster, G., G. McCray, P. Kariger, T. Dua, A. Titman and J. Chandna (2018). "Creation of the WHO Indicators of Infant and Young Child Development (IYCD): Metadata Synthesis across Ten Countries". Manuscript under review.

Law, N., D. Woo, J. de la Torre and G. Wong (2018). "A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2". UIS Information Paper No. 51. Montreal: UNESCO Institute for Statistics (UIS).

Lockheed, M., T. Prokic-Bruer and A. Shadrova (2015). *The Experience of Middle-Income Countries Participating in PISA 2000-2015*. Washington: World Bank; Paris: OECD.

Maddox, B. and L. Esposito (2011). 'Sufficiency Re-examined: A Capabilities Perspective on the Assessment of Functional Adult Literacy". *Journal of Development Studies,* Vol. 47, Issue 9, pp. 1315–1331.

Martin, M.O., I.V.S. Mullis and M. Hooper (Eds.) (2017). *Methods and Procedures in PIRLS 2016.* Boston: TIMSS and PIRLS International Study Center.

Martin, M.O., I.V.S. Mullis, P. Foy and M. Hooper (2016). *TIMSS 2015 International Results in Science*. Boston: TIMSS and PIRLS International Study Center.

McCoy, D.C., M, Black, B. Daelmans and T. Dua (2016). "Measuring Development in Children from Birth to Age 3 at Population Level". *Early Childhood Matters 2016*, pp. 34-39.

Miller, A.C., M.B. Murray, D.R. Thomson and M.C. Arbour (2016). "How Consistent are Associations between Stunting and Child Development? Evidence from a Meta-Analysis of Associations between Stunting and Multidimensional Child Development in Fifteen Low- and Middle-Income Countries". *Public Health Nutrition*, Vol. 19, Issue 8, pp. 1339-1347.

Mullis, I.V.S. and M.O. Martin (Eds.) (2017a). *TIMSS 2019 Assessment Frameworks*. Boston: TIMSS and PIRLS International Study Center.

Mullis, I.V.S. and M.O. Martin (Eds.) (2015). *PIRLS 2016 Assessment Framework (2nd)*. Boston: TIMSS and PIRLS International Study Center.

Mullis, I.V.S. and M.O. Martin (Eds.) (2014). *TIMSS Advanced 2015 Assessment Frameworks*. Boston: TIMSS and PIRLS International Study Center.

Mullis, I.V.S., M.O. Martin and T. Loveless (2016a). *20 Years of TIMSS. International Trends in Mathematics and Science Achievement, Curriculum and Instruction.* Boston: TIMSS and PIRLS International Study Center.

Mullis, I.V.S., M.O Martin, P. Foy and K.T. Drucker (2012). *PIRLS 2011 International Results in Reading*. Boston: TIMSS and PIRLS International Study Center.

Mullis, I.V.S., M.O. Martin, P. Foy and M. Hooper (2017b). *PIRLS 2016 International Results in Reading*. Boston: TIMSS and PIRLS International Study Center.

Mullis, I.V.S., M.O. Martin, P. Foy and M. Hooper (2016b). *TIMSS 2015 International Results in Mathematics*. Boston: TIMSS and PIRLS International Study Center.

Mullis, I.V.S., M.O. Martin, S. Goh and K. Cotter (Eds.) (2016c). *TIMSS 2015 Encyclopedia: Education Policy and Curriculum in Mathematics and Science*. Boston: TIMSS and PIRLS International Study Center.

Mullis, I.V.S., M.O. Martin, S. Goh and C. Prendergast (Eds.) (2017c). *PIRLS 2016 Encyclopedia: Education Policy and Curriculum in Reading*. Boston: TIMSS and PIRLS International Study Center.

National Council for Educational Research and Training (NCERT) (2015). "What Students of Class V Know and Can Do: A Summary of India's National Achievement Survey, Class V (Cycle 4), 2015". New Delhi: NCERT.

National University for Educational Planning and Administration (2017). "Elementary Education in India: Progress towards UEE". New Delhi: National University of Educational Planning and Administration.

Niece, D. and T.S. Murray (1997). "Literacy Skills and the Readiness of Adults for Lifelong Learning and Further Education and Training: Trends based on the International Adult Literacy Survey".

Organisation for Economic Co-operation and Development (OECD) (2018). *Education at a Glance 2018: OECD Indicators*. Paris: OECD Publishing.

OECD (2017). "Where did equity in education improve over the past decade?" PISA in Focus, No. 68. Paris: OECD.

OECD (2016a). *PISA 2015 Results (Volume I): Excellence and Equity in Education.* Paris: OECD.

OECD (2016b). PISA 2015 Database. http://www.oecd.org/pisa/data/2015database/ Accessed July 2018.

OECD (2016c). *Skills Matter: Further Results from the Survey of Adult Skills*. Paris: OECD.

OECD (2016d). *Survey of Adult Skills Technical Report (2nd)*. Paris: OECD.

OECD (2016e). *The Survey of Adult Skills: Reader's Companion (2nd)*. Paris: OECD.

OECD (2014). *PISA 2012 Results: What Students Know and Can Do.* Volume I, Revised edition, February 2014, "Student Performance in Mathematics, Reading and Science". Paris: OECD.

OECD (2013a). "What Makes Urban Schools Different?" *PISA in Focus*, No. 28. Paris: OECD.

OECD (2013b). *PISA 2012 Results: What Makes Schools Successful*. Volume IV, "Resources, Policies and Practices". Paris: OECD.

OECD (2013c). *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Paris: OECD.

OECD (2006). *PISA 2003 Technical Report*. Paris: OECD.

OECD/Statistics Canada (2011). *Literacy for Life: Further Results from the Adult Literacy and Life Skills Survey*. Paris: OECD.

OECD/Statistics Canada (2000). *Literacy in the Information Age: Final Report of the International Adult Literacy Survey*. Paris: OECD.

PASEC (2014). "Education System Performance in Francophone Sub-Saharan Africa: Competencies and Learning Factors in Primary Education". http://www.pasec.confemen.org/wp-content/uploads/2015/12/Rapport_Pasec2014_GB_webv2.pdf

Piper, B. and A. Mugenda (2014). "The Primary Math and Reading (PRIMR) Initiative: Endline Impact Evaluation". Research Triangle Park, NC: RTI International.

Piper, B., J. DeStefano, E. Kinyanjui and S. Ong'ele (2018). "Scaling up successfully: Lessons from Kenya's Tusome national literacy program". *Journal of Educational Change*, Vol. 19, Issue 3, pp. 293-321.

Platas, L. M., L.R. Ketterlin-Geller and Y. Sitabkhan (2016). "Using an Assessment of Early Mathematical Knowledge and Skills to inform Policy and Practice: Examples from the Early Grade Mathematics Assessment". *International Journal of Education in Mathematics, Science and Technology*, Vol. 4, Issue 3, pp. 163-173.

Platas, L.M., L.R. Ketterlin-Geller, A. Bombacher and Y. Sitabkhan (2014). “Early Grade Mathematics Assessment (EGMA) Toolkit”. Research Triangle Park, NC: RTI International.

Põldoja, H., T. Väljataga, M. Laanpere and K. Tammets (2014). Web-based Self- and Peer-Assessment of Teachers’ Digital Competencies. *World Wide Web*, Vol. 17, Issue 2, pp. 255-269.

Raikes, A., T. Dua and P.R. Britto (2015). “Measuring Early Childhood Development: Priorities for post-2015”. *Early Childhood Matters 2015*, pp. 74-77, The Hague, Netherlands: Bernard van Leer Foundation.

Raikes A., H. Yoshikawa, P.R. Britto and I. Iruka (2017). “Children, youth and developmental science in the 2015-2030 global sustainable development goals”. Society for Research in Child Development, Social Policy Report 30:(3).

Ramirez, M.J. (2018a). “How can countries monitor learning at the national level?”. Background paper for the 2018 SDG 4 Data Digest.

Ramirez, M.J. (2018b). “Challenges in communicating and using national learning assessments”. Background paper for the 2018 SDG 4 Data Digest.

Reckase, M. (2000). “Test Theory: A Unified Treatment”. *Applied Psychological* Measurement, Vol. 24, No. 2, June, pp. 187-189. Washington, D.C.: Sage Publications.

Rocher, T. (2015). “Mesure des compétences. Méthodes psychométriques utilisées dans le cadre des évaluations des élèves“. Éducation & Formation, Vol. 86-87, pp. 37-60.

RTI International (2016). “Early Grade Reading Assessment (EGRA) Toolkit (2nd)”. Washington: RTI International.

RTI International (2009). “Early Grade Mathematics Assessment (EGMA): A Conceptual Framework based on Mathematics Skills Development in Children”. Research Triangle Park, NC: RTI International.

RTI International (2007). “Early Grade Reading Assessment (EGRA) Workshop Notes.” Summary Notes from the Expert Workshop, November 16-17, 2006. Research Triangle Park, NC: RTI International.

Rutkowski, D. and L. Rutkowski (2018). “How Systemic is International Bullying and What Relationship Does It Have with Mathematics Achievement in 4th grade?” IEA Compass: Briefs in Education, No. 1. Amsterdam: IEA.

Schulz, W., J. Ainley, J. Fraillon, B. Losito and G. Agrusti (2016). *IEA International Civic and Citizenship Education Study 2016 Assessment Framework*. Cham: Springer.

Schwippert, K. (2003). *Progress in Reading Literacy: The Impact of PIRLS 2001 in 13 Countries*. Münster: Waxmann.

Schwippert, K. and J. Lenkeit (2012). *Progress in Reading Literacy in National and International Context: The Impact of PIRLS 2006 in 12 Countries*. Münster: Waxmann.

Slavin, R. E. (2002). “Evidence-Based Education Policies: Transforming Educational Practice and Research”. *Educational Researcher,* Vol. 31, Issue 7, pp. 15-21

Sparks, J. R., Katz, I. R., and Beile, P. M. (2016). “Assessing digital information literacy in higher education: A review of existing frameworks and assessments with recommendations for next-generation assessment”. Research Report No. RR-16-32. Princeton, NJ: Educational Testing Service.

Spaull, N. (2017). “Who Makes It Into PISA? Understanding the Impact of PISA Sample Eligibility Using Turkey as a Case Study (PISA 2003-PISA 2012)”. OECD Education Working Paper, No. 154. Paris: OECD.

Treviño, E. and M. Ordenes (2017). “Exploring Commonalities and Differences in Regional and International Assessments”. UIS Information Paper No. 48. Montreal: UNESCO Institute for Statistics (UIS).

United Nations (2015). *Principles and Recommendations for Population and Housing Censuses, Revision 3*. New York: United Nations.

UNESCO (2018). *Global Education Monitoring Report 2018. Accountability in Education: Meeting our Commitments* (2018). Paris: UNESCO.

UNESCO (2017a). *Cracking the Code: Girls' and Women's Education in Science, Technology, Engineering and Mathematics.* Paris: UNESCO.

UNESCO/IEA (2017). *Measuring SDG 4: How PIRLS Can Help*. Paris: UNESCO.

UNESCO Institute for Lifelong Learning (UIL) (2018). *Recherche-action sur la mesure des apprentissages des bénéficiaires des programmes d'alphabétisation. Le référentiel de compétences harmonisé*. Hamburg: UIL.

UNESCO Institute for Statistics (UIS) (2018a). "The Investment Case for SDG 4 Data". Technical Cooperation Group on SDG 4-Education 2030 Indicators. Montreal: UIS.

UNESCO Institute for Statistics (UIS) (2018b). "Towards an Innovative Demand-Driven Global Strategy for Education Data". Montreal: UIS.

UNESCO Institute for Statistics (UIS) (2018c). "A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2". Montreal: UIS.

UNESCO Institute for Statistics (UIS) (2017a). "SDG Data Reporting: Proposal of a Protocol for Reporting Indicator 4.1.1". GAML working paper. Montreal: UNESCO Institute for Statistics.

UNESCO Institute for Statistics (UIS) (2017b). *SDG4 Data Digest 2017 - The Quality Factor: Strengthening National Data to Monitor Sustainable Development Goal 4.* Montreal: UIS.

UNESCO Institute for Statistics (UIS) (2017c). *Quick Guide No. 2: Making the Case for a Learning Assessment.* Montreal: UIS.

UNESCO Institute for Statistics (UIS) (2017d). *Quick Guide No. 3: Implementing a National Learning Assessment.* Montreal: UIS.

UNESCO Institute for Statistics (UIS) (2017e). "Investment Case for Expanding Coverage and Comparability for Global Indicator 4.1.1". Montreal: UIS.

UNESCO Institute for Statistics (UIS) (2017f). "SDG Indicator 4.1.1: Inputs to the Measurement and Reporting Strategy". Montreal: UIS.

UNESCO Institute for Statistics (UIS) (2017g). "More Than One-Half of Children and Adolescents Are Not Learning Worldwide". UIS Fact Sheet No. 46. Montreal: UIS.

UNESCO Institute for Statistics (UIS) (2017h). "Measurement Strategy for SDG Target 4.7: Proposal by GAML Task Force 4.7". GAML Fourth Meeting. Montreal: UIS.

UNESCO Institute for Statistics (UIS) (2016). *Understanding What Works in Oral Reading Assessments: Recommendations from Donors, Implementers and Practitioners.* Montreal: UIS.

UNESCO Institute for Statistics (UIS) (2017i). "Review of the Use of Cross-National Assessment Data in Educational Practice and Policy". Montreal: UIS.

UNESCO Institute for Statistics (UIS) (2012). *ISCED 1997: International Standard Classification of Education*. Montreal: UIS.

UNESCO Institute for Statistics (UIS) and ACER (2017). "Principles of Good Practice in Learning Assessment". Montreal: UIS.

UNICEF (2018). "Multiple Indicator Cluster Survey". http://mics.unicef.org/

UNICEF (2017). *Early Moments Matter*. New York: UNICEF.

UN Statistics Division (2015). “Metadata: Target 4.6”. https://unstats.un.org/sdgs/metadata/files/Metadata-04-06-01.pdf

USAID (2018). “Early Grade Reading Barometer, Comparisons Page”. http://www.earlygradereadingbarometer.org/nigeria-jigawa/comparisons/sdg

Varone, F. (2007). “Développer les capacités d’évaluation. L’évaluation des politiques au niveau régional“. *Collection « Action publique »,* Vol. 1.

Verma, S. and A. Petersen (Eds.) (2018). *Developmental Science and Sustainable Development Goals for Children and Youth*. New York: Springer International Publishing.

Vuorikari, R., Y. Punie, S. Carretero and L. Van den Brande (2016). “DigComp 2.0: The Digital Competence Framework for Citizens Update Phase 1: The Conceptual Reference Model”. https://ec.europa.eu/jrc/en/DigComp/digital-competence-framework

Willms, J.D. (2018). “Learning Divides: Using Data to Inform Educational Policy”. UIS Information Paper No. 54. Montreal: UNESCO Institute for Statistics (UIS).

World Bank (2018). *The World Development Report 2019*. Washington: World Bank.

World Bank (2017a). *The World Development Report 2018*. Washington: World Bank.

World Bank (2017b). “World Bank Country and Lending Groups”. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

World Bank (2014). *STEP Skills Measurement Surveys: Innovative Tools for Assessing Skills*. Washington D.C.: World Bank.

Yoshikawa H., A. Raikes and A. Wuermli (2017), “Measurement Options for Development of Sustainable Development Goal Indicator 4.2.1”. Global Alliance to Monitor Learning, Task Force on Target 4.2. Montreal: UNESCO Institute for Statistics (UIS).

Young Lives (2014). “Education and Learning: Round 4 Preliminary Findings”. India: Young Lives.

# Further Readings


Altinok, N., N. Angrist and H.A. Patrinos (2018). "Global Dataset on Education Quality (1965-2015)". Washington: World Bank.

Bond, T. and C. Fox (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences (2nd).* Mahwah, NJ: Lawrence Erlbaum Associates.

Crouch, L., and K.A. Merseth (2017). "Stumbling at the First Step: Efficiency Implications of Poor Performance in the Foundational Five Years. *Quarterly Review of Comparative Education*, Vol. 47, Issue 3, pp. 175-196.

Filmer, D. and L. Pritchett (2001). "Estimating Wealth Effects without Expenditure Data–or Tears: An Application to Educational Enrolments in States of India". *Demography*, Vol. 38, Issue 1, pp. 115-132.

García, J.L., J.J. Heckman, D.E. Leaf and M.J. Prados (2016). "The Life-Cycle Benefits of an Influential Early Childhood Program". Human Capital and Economic Opportunity Global Working Group, Working Paper 2016-035. Chicago: Heckman.

IEA (2018). "The Data Makes the Difference: How Chinese Taipei used TIMSS Data to Reform Mathematics Education". IEA Compass: Briefs in Education, No. 2. Amsterdam: IEA.

Loizillon, A., N. Petrowsk, P. Britto and C. Cappa (2017). "Development of the Early Childhood Development Index in MICS Surveys". MICS Methodological Papers, No. 6. New York: UNICEF.

Richter, L. M., B. Dealmans and G.L. Darmstadt (2016). "Investing in the Foundation of Sustainable Development: Pathways to scale up for Early Childhood Development". *The Lancet*, Vol. 389(*10064*), pp. 103–118.

Schulz, W., J. Ainley, J. Fraillon, B. Losito and G. Agrusti (2016). *IEA International Civic and Citizenship Education Study 2016 Assessment Framework*. Cham: Springer.

Senghor, O. (2014). "Disseminating and using student assessment information in the Gambia". SABER-Student Assessment Working Paper No. 11. Washington: World Bank.

Shonkoff, J.P., A.S. Garner, the Committee on Psychosocial Aspects of Child and Family Health, the Committee on Early Childhood, Adoption and Dependent Care, and the Section on Behavioural Pediatrics (2012). "The Lifelong Effects of Early Childhood Adversity and Toxic Stress". *Pediatrics*, Vol. 129, Issue 1, pp. 232-246.

UNESCO Institute for Statistics (UIS) (2006). *ISCED 1997: International Standard Classification of Education*. Montreal: UIS.

Wamani, H., T. Tylleskär, A.N. Åstrøm, J.K. Tumwine and S. Peterson (2004). "Mothers' education but not fathers' education, household assets or land ownership is the best predictor of child health inequalities in rural Uganda". *International Journal for Equity in Health*, Vol.3 Issue 1, pp. 1-8.

Zheng, X. and S. Rabe-Hesketh (2007). "Estimating Parameters of Dichotomous and Ordinal Item Response Models with GLAMM". *Stata Journal*, Vol. 7, Issue 3, pp. 313-333.

# Annex 1. List of global and thematic indicators

| Indicator number | Indicator description | Global/ thematic | Custodial agency |
|---|---|---|---|
| Target 4.1. | By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes | | |
| 4.1.1 | Proportion of children and young people (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex | Global | UIS |
| 4.1.2 | Administration of a nationally-representative learning assessment (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education | Thematic | |
| 4.1.3 | Gross intake ratio to the last grade (primary education, lower secondary education) | | |
| 4.1.4 | Completion rate (primary education, lower secondary education, upper secondary education) | | |
| 4.1.5 | Out-of-school rate (primary education, lower secondary education, upper secondary education) | | |
| 4.1.6 | Percentage of children over-age for grade (primary education, lower secondary education) | | |
| 4.1.7 | Number of years of (a) free and (b) compulsory primary and secondary education guaranteed in legal frameworks | | |
| Target 4.2 | By 2030, ensure that all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education | | |
| 4.2.1 | Proportion of children under 5 years of age who are developmentally on track in health, learning and psychosocial well-being, by sex | Global | UNICEF |
| 4.2.2 | Participation rate in organized learning (one year before the official primary entry age), by sex | Global | UIS |
| 4.2.3 | Percentage of children under 5 years experiencing positive and stimulating home learning environments | Thematic | |
| 4.2.4 | Gross early childhood education enrolment ratio in (a) pre-primary education and (b) early childhood educational development | | |
| 4.2.5 | Number of years of (a) free and (b) compulsory pre-primary education guaranteed in legal frameworks | | |
| Target 4.3. | By 2030, ensure equal access for all women and men to affordable and quality technical, vocational and tertiary education, including university | | |
| 4.3.1 | Participation rate of youth and adults in formal and non-formal education and training in the previous 12 months, by sex | Global | UIS |

| Indicator number | Indicator description | Global/ thematic | Custodial agency |
|---|---|---|---|
| 4.3.2 | Gross enrolment ratio for tertiary education by sex | Thematic | |
| 4.3.3 | Participation rate in technical-vocational programmes (15- to 24-year-olds) by sex | | |
| Target 4.4. | By 2030, substantially increase the number of youth and adults who have relevant skills, including technical and vocational skills, for employment, decent jobs and entrepreneurship | | |
| 4.4.1 | Proportion of youth and adults with information and communications technology (ICT) skills, by type of skill | Global | UIS/ITU |
| 4.4.2 | Percentage of youth/adults who have achieved at least a minimum level of proficiency in digital literacy skills | Thematic | |
| 4.4.3 | Youth/adult educational attainment rates by age group, economic activity status, levels of education and programme orientation | | |
| Target 4.5 | By 2030, eliminate gender disparities in education and ensure equal access to all levels of education and vocational training for the vulnerable, including persons with disabilities, indigenous peoples and children in vulnerable situations | | |
| 4.5.1 | Parity indices (female/male, rural/urban, bottom/top wealth quintiles and others such as disability status, indigenous peoples and conflict-affected, as data become available) for all education indicators on this list that can be disaggregated | Global | UIS |
| 4.5.2 | Percentage of students in primary education whose first or home language is the language of instruction | Thematic | |
| 4.5.3 | Extent to which explicit formula-based policies reallocate education resources to disadvantaged populations | | |
| 4.5.4 | Education expenditure per student by level of education and source of funding | | |
| 4.5.5 | Percentage of total aid to education allocated to least developed countries | | |
| Target 4.6 | By 2030, ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy | | |
| 4.6.1 | Proportion of population in a given age group achieving at least a fixed level of proficiency in functional (a) literacy and (b) numeracy skills, by sex | Global | UIS |
| 4.6.2 | Youth/adult literacy rate | Thematic | |
| 4.6.3 | Participation rate of illiterate youth/adults in literacy programmes | | |
| Target 4.7. | By 2030, ensure that all learners acquire the knowledge and skills needed to promote sustainable development, including, among others, through education for sustainable development and sustainable lifestyles, human rights, gender equality, promotion of a culture of peace and non-violence, global citizenship and appreciation of cultural diversity and of culture's contribution to sustainable development | | |

| Indicator number | Indicator description | Global/ thematic | Custodial agency |
|---|---|---|---|
| 4.7.1 | Extent to which (i) global citizenship education and (ii) education for sustainable development, including gender equality and human rights, are mainstreamed at all levels in: (a) national education policies, (b) curricula, (c) teacher education and (d) student assessment | Global | UIS |
| 4.7.2 | Percentage of schools that provide life skills-based HIV and sexuality education | Thematic | |
| 4.7.3 | Extent to which the framework on the World Programme on Human Rights Education is implemented nationally (as per the UNGA Resolution 59/113) | | |
| 4.7.4 | Percentage of students by age group (or education level) showing adequate understanding of issues relating to global citizenship and sustainability | | |
| 4.7.5 | Percentage of 15-year-old students showing proficiency in knowledge of environmental science and geoscience | | |
| Target 4.a. | Build and upgrade education facilities that are child, disability and gender sensitive and provide safe, non-violent, inclusive and effective learning environments for all | | |
| 4.a.1 | Proportion of schools with access to: (a) electricity; (b) Internet for pedagogical purposes; and (c) computers for pedagogical purposes | Global | UIS |
| 4.a.2 | Percentage of students experiencing bullying | Thematic | |
| 4.a.3 | Number of attacks on students, personnel and institutions | | |
| Target 4.b. | By 2020, substantially expand globally the number of scholarships available to developing countries, in particular least developed countries, small island developing States and African countries, for enrolment in higher education, including vocational training and information and communications technology, technical, engineering and scientific programmes, in developed countries and other developing countries | | |
| 4.b.1 | Volume of official development assistance flows for scholarships by sector and type of study | Global | OECD |
| 4.b.2 | Number of higher education scholarships awarded by beneficiary country | Thematic | |
| Target 4.c. | By 2030, substantially increase the supply of qualified teachers, including through international cooperation for teacher training in developing countries, especially least developed countries and small island developing States | | |
| 4.c.1 | Proportion of teachers in: (a) pre-primary education; (b) primary education; (c) lower secondary education; and (d) upper secondary education who have received at least the minimum organized teacher training (e.g., pedagogical training) pre-service or in-service required for teaching at the relevant level in a given country, by sex | Global | UIS |

| Indicator number | Indicator description | Global/ thematic | Custodial agency |
|---|---|---|---|
| 4.c.2 | Pupil-trained teacher ratio by education level | Thematic | |
| 4.c.3 | Percentage of teachers qualified according to national standards by education level and type of institution | | |
| 4.c.4 | Pupil-qualified teacher ratio by education level | | |
| 4.c.5 | Average teacher salary relative to other professions requiring a comparable level of qualification | | |
| 4.c.6 | Teacher attrition rate by education level | | |
| 4.c.7 | Percentage of teachers who received in-service training in the last 12 months by type of training | | |

*Source: UNESCO Institute for Statistics (UIS).*

# Annex 2. IEA's Rosetta Stone: Measuring global progress toward the SDG for quality education by linking regional assessment results to TIMSS and PIRLS international benchmarks of achievement*

This IEA proposal is to address the need to measure progress toward the UN SDG 4: Ensure inclusive and quality education for all and promote lifelong learning. In particular, the proposal describes a strategy for developing Indicator 4.1.1: Proportion of children and young people: (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex. As set forth in "Unpacking Sustainable Development Goal 4: Education 2030: Guide", the principles, strategies and actions for this target go beyond the simple dichotomy of "literate" versus "illiterate" and are underpinned by the contemporary understanding of literacy as a continuum of proficiency levels. More specifically, the guide states that "action for this target aims at ensuring that by 2030 all young people and adults across the world should have achieved relevant and recognised proficiency levels in functional literacy and numeracy skills".

IEA's TIMSS and PIRLS international assessments provide widely-recognised proficiency levels in numeracy and literacy, respectively, for students at the end of primary schooling. TIMSS has been measuring trends in mathematics and science at four-year intervals since 1995. PIRLS has measured trends in reading literacy at five-year intervals since 2001. With 50 to 70 countries participating in each assessment cycle, the TIMSS and PIRLS achievement scales and their International Benchmarks are well established and used by countries all around the world. Especially pertinent to measuring progress to the SDG goals, both TIMSS and PIRLS have devoted considerable resources to extending their achievement scales to provide high quality measurement for countries where most children still are developing basic numeracy and literacy skills. For example, the PIRLS assessment has been doubled in scope with the same amount of coverage allocated to a less difficult version of PIRLS that assesses literacy with shorter and simpler texts. It also has reading passages in common with PIRLS such that students can participate primarily with literacy passages and items and still be reported on the PIRLS achievement scale. Similarly, TIMSS mathematics now includes a less difficult assessment providing a comprehensive measurement of basic numeracy skills.

* Written by the International Evaluation Association (IEA).

### A2.1 Objective

The proposal presents a strategy for providing information about the proportions of primary school students that have achieved established proficiency levels in literacy and numeracy. The aim is to establish a link between the results on regional assessments conducted at the primary level and the TIMSS and PIRLS International Benchmarks for numeracy and literacy. There are five regional assessments planning reading and mathematics assessments at the end of primary schooling in 2018 or 2019:

- SACMEQ
- PASEC
- LLECE
- SEA-PLM
- PILNA

The reading and mathematics assessments planned for 2018/2019 provide a perfect opportunity to link these regional assessment results to IEA's TIMSS and PIRLS achievement scales. These regional assessments measure achievement at the sixth grade, except SEA-PLM which is at the fifth grade. The content of the regional mathematics assessments align well with the TIMSS fourth grade assessments of numeracy and mathematics. Similarly, the content of the regional reading assessments align well with the PIRLS fourth grade assessment of literacy and reading comprehension. The overarching idea is to construct a concordance table that translates between the scores on each of the regional assessments in mathematics and reading and scores on TIMSS and PIRLS, respectively.

The concordance table is the "Rosetta Stone" that provides a translation from the countries' regional assessment results to the TIMSS and PIRLS achievement scales. Similar to the original Rosetta Stone, which provided a link between Greek and Egyptian hieroglyphics, the concordance table provides a link between regional assessments and the TIMSS and PIRLS achievement scales. The countries participating in the regional assessments can use the translations to determine what percentage of their students could be expected to reach the TIMSS and PIRLS International Benchmarks.

### A2.2 Implementation

The IEA will work with the study centres for each of the five regional assessments. The proposal is to have a sub-set of countries (3 to 5) from each regional assessment administer selected booklets of TIMSS and PIRLS achievement items at the same time as their upcoming regional assessments. Depending on the level of mathematics and reading achievement in a region, the booklets can be tailored to contain primarily items assessing TIMSS Numeracy and PIRLS Literacy. The same students should take the regional mathematics and reading assessments and then also the TIMSS and PIRLS booklets, preferably on the following day. The combined data across the three to five countries will provide scores on both the regional assessment and TIMSS and PIRLS for approximately 15,000 students from the region that can be used to construct the "Rosetta Stone" concordance tables for numeracy and literacy achievement. For each regional assessment, because the concordance tables provide a projected TIMSS or PIRLS score for all possible regional assessment scores, it will be possible to determine the regional assessment scores equivalent to each of the TIMSS and PIRLS International Benchmarks.

TIMSS and PIRLS each have four international benchmarks – Low (400), Intermediate (475), High (550) and Advanced (625). For each country participating in a regional assessment, progress toward an international benchmark can be estimated by the percentage of students reaching the regional assessment score equivalent to the international benchmark. For example, a country may want to determine the percentage of students reaching the 'Low' international benchmark. Hypothetically, if the concordance table showed that a regional assessment score of 562 in reading was equivalent to 400 on the PIRLS reading scale, then all students in the country reaching 562 could be considered to have

reached the 'Low' international benchmark. Although based on data from the three to five countries that participate in the linking study, the concordance table and the benchmark equivalent scores can be applied in all the countries in the regional assessment (whether they participated in the linking study or not).

### A2.3 Schedule

The Rosetta Stone Linking Project for regional assessments will take four years: 2018 to 2021.

2018: Meet with regional study centres to plan operations; prepare TIMSS and PIRLS assessment booklets and data collection manuals.

2019: Conduct linking data collection in accordance with regional assessment schedules; conduct training in constructed response item scoring.

2020: Prepare for and conduct psychometric scaling of regional assessment and TIMSS and PIRLS data and construct concordance tables.

2021: Produce the reports to regional assessment study centers, including technical documentation about the match between the assessment frameworks and assessment items for the regional assessments and TIMSS and PIRLS and the methodology employed.

# Annex 3. Social moderation method for linking national and cross-national assessments to the UIS Proficiency Scale*

The purpose of this paper is to explain, to a broad audience, how results from national and cross-national student assessments (NAs and CNAs) can be put on the same scales in order to report on Indicator 4.1.1. The UIS, as the custodial agency for reporting on SDG 4, needs to develop these scales that support national governments in effectively measuring and monitoring student learning outcomes to report against Indicator 4.1.1 over time. The reporting will take place at three education levels – Grade 2 or 3, at the end of primary education and at the end of lower secondary education – in two subject areas – reading and mathematics. This will require a total of six scales for reporting.

There are three possibilities for assessing students for UIS reporting. First, the UIS or another agency could develop and administer a common assessment in all countries. This has been discussed but deemed not feasible due to time, cost and consensus-building requirements.

Second, the UIS could fund statistical linking of NAs and CNAs with a UIS scale. This would require using test- or item-based linking methods (i.e. equating, calibration, projection or statistical moderation) by embedding anchor or common items in each of these assessments (common item equating) or having the same students take multiple assessments (common person equating). GAML and a technical partner, the Australian Council for Educational Research (ACER), are exploring the development of reporting scales – the UIS Reporting Scale – that would facilitate statistical linking of NAs and CNAs using anchor or common items. There is broad support for exploring the development of the UIS Reporting Scale and conducting statistical linking, but there is also recognition that it is a long-term effort, with cost and possible test security issues. Another promising effort in statistical linking involves embedding anchor or common items into either TIMSS or PIRLS and the corresponding regional assessments, but this is a long-term effort with security issues as well.

Third, in order to satisfy the more immediate need for UIS reporting on Indicator 4.1.1, the authors are proposing a process involving a method called social moderation or policy linking. This is a non-statistical linking procedure that uses definitions of proficiency levels for reading and mathematics to produce a reporting scale – called a proficiency scale, in this instance – and a mechanism for linking existing assessments and their performance levels to this scale. This could take place relatively quickly. Several steps are involved in constructing proficiency scales – the UIS Proficiency Scale – that would facilitate the non-statistical linking of NAs and CNAs to that scale.

In brief, six steps, with related outputs identified below, are involved:

1. *Define content standards:* what students are expected to learn in reading and mathematics at the three education levels, i.e. Grade 2 or 3, at the

* Written by Dana Kelly, Jeff Davis and Abdullah Ferdous, Technical Diretor, Managemente Systems Internaitonal (MSI).

end of primary education and at the end of lower secondary education.
2. *Determine performance levels:* number of categories and names for levels (e.g. levels separated by minimum proficiency in reading and mathematics at each education level).
3. *Develop policy definitions of performance (PDPs):* what students should demonstrate in each category, in generic terms and not by subject area, at each education level.
4. *Develop performance level descriptors (PLDs):* what students should demonstrate in reading and mathematics (details of knowledge, skills, abilities) at each education level.
5. *Develop proficiency scale maps:* how performance levels of various NAs and CNAs map to the UIS Proficiency Scale in reading and mathematics at each education level.
6. *Develop socially-moderated performance standards:* what students need to score on NAs and CNAs in reading and mathematics at each education level for placement into categories.

Outputs 1 to 4 facilitate constructing a UIS Proficiency Scale, and outputs 5 and 6 facilitate linking the UIS Proficiency Scale with NAs and CNAs. The steps and outputs are described below.

### Steps

The following six steps, with outputs, are proposed to construct and apply the UIS Proficiency Scale:

*Step 1: Define content standards.* In order to develop stand-alone reporting scales for each of the three education levels in reading and mathematics (i.e. six scales), the first step is to define the content standards for each domain and for each grade span of K-3, 4-6 and 7-9 separately. As mentioned above, the common content standards are predefined knowledge and skills that students are expected to learn in reading and mathematics by the end of Grades 3, 6 and 9 across countries. The UNESCO's International Bureau of Education (IBE-UNESCO) has made significant progress in describing these content standards for each domain and grade. It has already reviewed and analysed over 140 NAs and CNAs to identify the content standards of various grades being assessed (IBE and UIS, 2018; IBE and UIS, 2017).

*Step 2: Determine performance levels.* In this step, the number of levels to be used and their names on the scales are determined. Typically, no more than four performance levels are needed (Perie, 2008). Beyond four levels, it becomes difficult to describe meaningful differences across the levels. Three is probably advisable for the UIS Proficiency Scale, with a level below minimum proficiency and two levels of proficiency. After determining the number of levels, the next task is to name the levels. There are no clear-cut guidelines on how to develop names for the levels, however it is recommended that they be thoughtfully chosen to relate to the purpose of reporting and supportable inferences arising from the classifications (Cizek and Bunch, 2007).

*Step 3: Develop policy definitions of performance (PDPs).* The next step is to develop a generic policy definition for each performance level. These definitions are not linked to content but are more general statements that assert policymakers' position on the desired level of performance. They are particularly useful in the context of reporting multiple assessments. First, they facilitate the articulation of performance levels across grades by ensuring the same level of rigor at each level across each grade. Second, they allow a reader to interpret proficiency in a similar manner regardless of subject assessed. The policy definitions need to be written for each level. **Figure A3.1** presents an example from a national assessment programme (U.S. National Assessment of Educational Progress (NAEP)), with three categories: basic, proficient and advanced.

In writing policy definitions for performance levels, it is strongly recommended that they distinguish clearly among the levels. The definitions should state the degree of knowledge and skills expected of students at each level. They should be concise, approximately one to two sentences, and clear (Perie, 2008).

**Figure A3.1 Illustrative policy definitions of performance (PDPs)**

| Performance levels | | |
|---|---|---|
| **Basic:** This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade. | **Proficient:** Solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter. | **Advanced:** This level signifies superior performance beyond proficient. |

*Source: Authors.*

Because the PDP is the backbone of full descriptions (in the next step), the UIS should carefully consider the wording and be sure each definition communicates the intended goals and clearly distinguishes one level from the next.

*Step 4: Develop performance level descriptors (PLDs).* After the policy definitions have been adopted, full performance level descriptors (PLDs) should be developed for each education level and each subject area (reading and mathematics). The full descriptions express the knowledge and skills required to achieve the performance levels. They can be used to provide stakeholders with more information on what students at each performance level should know and be able to do, as well as what they need to know and be able to do to reach the next performance level.

To develop full descriptions, for each domain a PLD writing workshop is conducted with subject matter experts. Five to eight people per subject and grade span will suffice (Perie, 2008). The subject matter experts will start with the policy definitions and expand those definitions in terms of specific knowledge, skills and abilities at each education level for each domain. The PLDs should be very detailed and reflect the content standards defined in Step 1. **Figure A3.2** provides an example of PLDs for reading at the end of primary education, adapted from a U.S. statewide assessment programme, with the same three levels used in the PDPs.

Since the PLDs of the UIS Proficiency Scale will be the basis for linking with NAs and CNAs, it is essential that they are fully elaborated and include details related to each content standard identified in Step 1.

*Step 5: Develop proficiency scale maps.* After performance levels of the UIS Proficiency Scale for each grade and domain are determined and described, the next step is to link the UIS Proficiency Scale for each education level and subject area with corresponding NAs and CNAs for Indicator 4.1.1 reporting. The different assessments can be linked through the PLDs. This process is called social moderation or policy linking (Reckase, 2000).

In order to explain the social moderation process, let us assume that the UIS Proficiency Scale for the end of primary education reading has four performance levels or categories. Levels 1 and 2 are sub-levels within 'Basic' or 'Below minimum proficiency', with Level 2 corresponding to 'Proficient' and 'Exceeds to advanced'.

1. *Does not meet minimum proficiency*
2. *Partially meets minimum proficiency*
3. *Meets minimum proficiency*
4. *Exceeds minimum proficiency*

**Figure A3.2 Illustrative performance level descriptors (PLDs)**

| Performance levels | | |
|---|---|---|
| **Basic:** A student performing at this level demonstrates limited comprehension of literary and informational texts and may use textual evidence to summarise and/or analyse a text. The student inconsistently analyses how an element of literature or informational text develops and influences the text. The student may determine a central idea in an informational text. The student may determine how the author uses organization, structure, form, text features, figurative language, and/or word choice to achieve a purpose. The student determines the point of view in a text. The student provides an incomplete comparison between texts in different forms or genres. The student may identify the development of an argument and may evaluate the author's claims and evidence in a text. The student may use context and word structure to determine the meanings of words, may interpret figurative language, and may understand some word meanings. | **Proficient**: A student performing at this level demonstrates comprehension of literary and informational texts by using textual evidence to summarise and/or analyse a text. The student analyses how an element of literature or informational text develops and influences the text. The student determines a central idea in an informational text. The student determines how the author uses organization, structure, form, text features, figurative language, and/or word choice to achieve a purpose. The student determines the effectiveness of point of view in a text. The student compares and contrasts texts in different forms or genres. The student traces the development of an argument and evaluates the author's claims and evidence in a text. The student uses context and word structure to determine the meanings of words, interprets figurative language, and understands nuances in word meanings. | **Advanced:** A student performing at this level demonstrates thorough comprehension of literary and informational texts by using key textual evidence to effectively summarise and/or analyse a text. The student thoroughly analyses how an element of literature or informational text develops and influences the text. The student determines a central idea in an informational text. The student determines how the author uses organization, structure, form, text features, figurative language, and/or word choice to achieve a purpose. The student determines the effectiveness of point of view in a text. The student thoroughly compares and contrasts texts in different forms or genres. The student traces the development of an argument and thoroughly evaluates the author's claims and evidence in a text. The student uses context and word structure to determine the meanings of words, interprets figurative language and understands nuances in word meanings. |

*Source: Authors.*

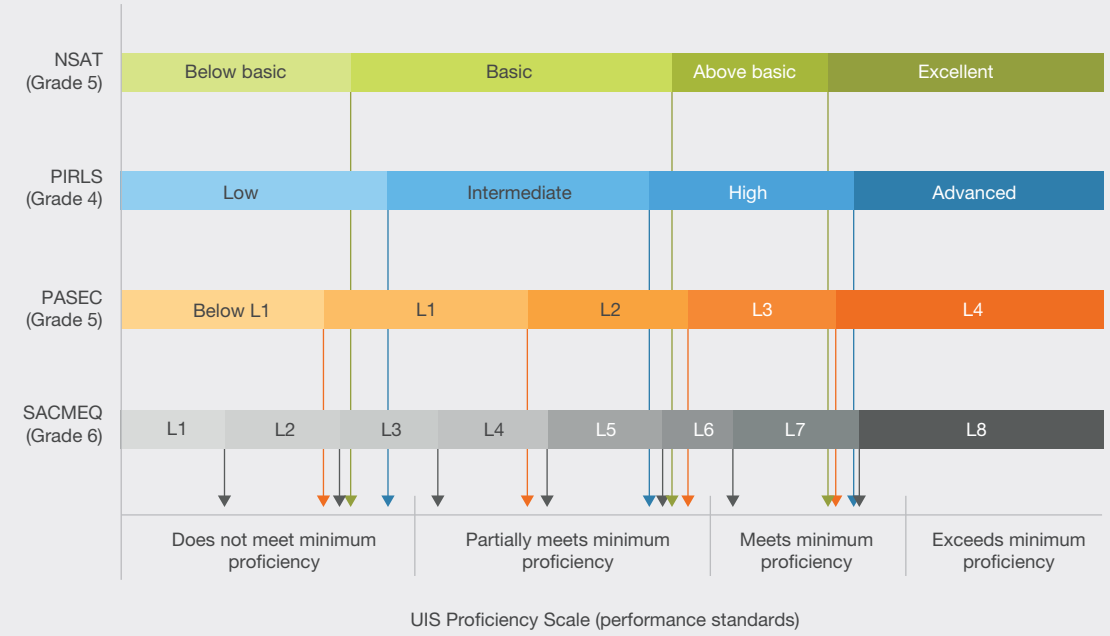
We have selected four levels for illustrative purposes only, since many of the international assessments have four levels rather than three levels that are suggested for the UIS Proficiency Scale. Note that the bottom two levels could be combined into a 'Basic' (or below minimum proficiency) category, as used in the PDPs and PLDs above. We can assume that the levels have been defined by both policymakers (i.e. PDPs) and other stakeholders (i.e. PLDs) according to Steps 3 and 4.

**Figure A3.3** shows examples of a proficiency scale map. It links one national assessment (Namibia's National Standardised Achievement Test or NSAT), one global assessment (PIRLS), and two regional assessments (PASEC and SACMEQ) with the UIS Proficiency Scale for reading at the end of primary education.

The map shows the degree to which the NSAT, PIRLS, PASEC and SACMEQ cut scores, i.e. the points at which the categories are separated for

the different assessments and line up with the cut scores of the UIS Proficiency Scale. As an example, the PIRLS middle (or proficient) cut score lines up the least well with the UIS Proficiency Scale middle (or proficient) cut score, with the PASEC cut score lining up the best. (Note that one of the reasons for this could be the different grade levels of the assessments.) The idea is then to determine new cut scores for the assessments, if necessary, so that they line up precisely with the cut scores of the UIS Proficiency Scale. By doing this, we will have alignment between the cut scores of the assessments and the cut scores of the UIS Proficiency Scale. The subject matter experts will make consensus ratings on the matches between performance levels of the NAs and CNAs with those of the UIS Proficiency Scale through social moderation procedures.

**Figure A3.3 Illustrative proficiency scale map with the UIS Proficiency Scale and selected assessments**




*Source: MSI.*

# Annex 4. Mapping of learning assessment data sources

## Global and thematic indicators map of existing cross-national learning initiatives

| Indicator number | Indicator description | Type of assessment | Major learning assessments collecting needed data | Questionnaire |
|---|---|---|---|---|
| Target 4.1. | By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes | | | |
| 4.1.1 | Proportion of children and young people (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex | School-based | EGMA/EGRA, PASEC, PILNA, PIRLS, PISA, SACMEQ, TERCE, TIMSS | Cognitive test |
| | | Household-based | PAL Network | Cognitive test |
| 4.1.3 | Gross intake ratio to the last grade (primary education, lower secondary education) | School-based | ICCS, ICILS, PIRLS, SACMEQ, TIMSS | School, principal |
| | | Household-based | MICS, PAL Network, STEP, Young Lives | School, household, individual |
| 4.1.4 | Completion rate (primary education, lower secondary education, upper secondary education) | Household-based | MICS, PAL Network, PIAAC, STEP, Young Lives | Household, individual |
| 4.1.5 | Out-of-school rate (primary education, lower secondary education, upper secondary education) | Household-based | MICS, PAL Network, PIAAC, STEP, Young Lives | Household, individual |
| 4.1.6 | Percentage of children over-age for grade (primary education, lower secondary education) | School-based | PASEC, PISA, SACMEQ, TERCE, TIMSS | Student |
| | | Household-based | MICS, PAL Network, STEP, Young Lives | Household, individual |
| 4.1.7 | Number of years of (a) free and (b) compulsory primary and secondary education guaranteed in legal frameworks | School-based | PIRLS, TIMSS | National context survey, principal, curriculum, school |
| | | Household-based | EAPECDScales, EHCI, IDELA, MELQO, MICS, Young Lives | Household |

| Indicator number | Indicator description | Type of assessment | Major learning assessments collecting needed data | Questionnaire |
|---|---|---|---|---|
| Target 4.2 | By 2030, ensure that all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education | | | |
| 4.2.1 | Proportion of children under 5 years of age who are developmentally on track in health, learning and psychosocial well-being, by sex | School-based | EDI, PIRLS, TIMSS | Teacher, school, home |
| | | Household-based | EAP ECD Scales, EHCI, IDELA, MELQO, MICS, PRIDI, Young Lives | Cognitive test, parent, teacher, individual |
| 4.2.2 | Participation rate in organized learning (one year before the official primary entry age), by sex | School-based | EDI, EGMA/EGRA, PASEC, PIRLS, PISA, SACMEQ, TERCE, TIMSS | Teacher, student, home |
| | | Household-based | EAP ECD Scales, EHCI, IDELA, MELQO, MICS, PAL Network, STEP, Young Lives | Parent, household, individual |
| 4.2.3 | Percentage of children under 5 years experiencing positive and stimulating home learning environments | School-based | PIRLS, TIMSS | Home |
| | | Household-based | EAPECDScales, EHCI, IDELA, MELQO, MICS, Young Lives | Parent, household, individual |
| 4.2.4 | Gross early childhood education enrolment ratio in (a) pre-primary education and (b) early childhood educational development | School-based | EDI, EGMA/EGRA, PASEC, PIRLS, PISA, SACMEQ, TERCE, TIMSS | Teacher, student, home |
| | | Household-based | EAP ECD Scales, EHCI, IDELA, MELQO, MICS, PAL Network, STEP, Young Lives | Parent, household, individual |
| 4.2.5 | Number of years of (a) free and (b) compulsory pre-primary education guaranteed in legal frameworks | School-based | ICCS | National context survey |
| | | Household-based | EAP ECD Scales, MICS, Young Lives | Parent, household |
| Target 4.3. | By 2030, ensure equal access for all women and men to affordable and quality technical, vocational and tertiary education, including university | | | |
| 4.3.1 | Participation rate of youth and adults in formal and non-formal education and training in the previous 12 months, by sex | Household-based | MICS, PAL Network, PIAAC, STEP, Young Lives | Household, individual |
| 4.3.2 | Gross enrolment ratio for tertiary education by sex | Household-based | MICS, PIAAC, STEP, Young Lives | Household, individual |
| 4.3.3 | Participation rate in technical-vocational programmes (15- to 24-year-olds) by sex | Household-based | MICS, PIAAC, STEP, Young Lives | Household, individual |

| Indicator number | Indicator description | Type of assessment | Major learning assessments collecting needed data | Questionnaire |
|---|---|---|---|---|
| Target 4.4. | By 2030, substantially increase the number of youth and adults who have relevant skills, including technical and vocational skills, for employment, decent jobs and entrepreneurship | | | |
| 4.4.1 | Proportion of youth and adults with information and communications technology (ICT) skills, by type of skill | School-based | ICILS | Student |
| | | Household-based | ITU, MICS, PIAAC, STEP | Household, individual |
| 4.4.2 | Percentage of youth/adults who have achieved at least a minimum level of proficiency in digital literacy skills | School-based | ICILS | Cognitive test |
| | | Household-based | ITU | Cognitive test |
| 4.4.3 | Youth/adult educational attainment rates by age group, economic activity status, levels of education and programme orientation | Household-based | MICS, PIAAC, STEP, Young Lives | Household, individual |
| Target 4.5 | By 2030, eliminate gender disparities in education and ensure equal access to all levels of education and vocational training for the vulnerable, including persons with disabilities, indigenous peoples and children in vulnerable situations | | | |
| 4.5.2 | Percentage of students in primary education whose first or home language is the language of instruction | School-based | EGMA/EGRA, PASEC, PIRLS, SACMEQ, TERCE, TIMSS | Student, principal, teacher, school, home |
| | | Household-based | MICS, PAL Network, STEP, Young Lives | Household, individual |
| 4.5.4 | Education expenditure per student by level of education and source of funding | School-based | PASEC, PISA | Principal, school |
| | | Household-based | EAP ECD Scales, PAL Network, Young Lives | parent, household, community |
| Target 4.6 | By 2030, ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy | | | |
| 4.6.1 | Proportion of population in a given age group achieving at least a fixed level of proficiency in functional (a) literacy and (b) numeracy skills, by sex | Household-based | PIAAC, STEP | Cognitive test |
| 4.6.2 | Youth/adult literacy rate | Household-based | MICS, PAL Network, PIAAC, STEP | Household, individual |
| 4.6.3 | Participation rate of illiterate youth/adults in literacy programmes | Household-based | STEP | Individual |

| Indicator number | Indicator description | Type of assessment | Major learning assessments collecting needed data | Questionnaire |
|---|---|---|---|---|
| Target 4.7. | By 2030, ensure that all learners acquire the knowledge and skills needed to promote sustainable development, including, among others, through education for sustainable development and sustainable lifestyles, human rights, gender equality, promotion of a culture of peace and non-violence, global citizenship and appreciation of cultural diversity and of culture's contribution to sustainable development | | | |
| 4.7.1 | Extent to which (i) global citizenship education and (ii) education for sustainable development, including gender equality and human rights, are mainstreamed at all levels in: (a) national education policies, (b) curricula, (c) teacher education and (d) student assessment | School-based | ICCS | National context survey |
| 4.7.2 | Percentage of schools that provide life skills-based HIV and sexuality education | School-based | SACMEQ | Student, principal, school, teacher |
| 4.7.4 | Percentage of students by age group (or education level) showing adequate understanding of issues relating to global citizenship and sustainability | School-based | ICCS | Cognitive test |
| 4.7.5 | Percentage of 15-year-old students showing proficiency in knowledge of environmental science and geoscience | School-based | PISA | Cognitive test |
| Target 4.a. | Build and upgrade education facilities that are child, disability and gender sensitive and provide safe, non-violent, inclusive and effective learning environments for all | | | |
| 4.a.1 | Proportion of schools with access to: (a) electricity; (b) Internet for pedagogical purposes; and (c) computers for pedagogical purposes | School-based | EGMA/EGRA, ICCS, ICILS, PASEC, PIRLS, PISA, SACMEQ, TERCE, TIMSS | Principal, school, ICT coordinator, teacher |
| | | Household-based | PAL Network, Young Lives | School, community |
| 4.a.2 | Percentage of students experiencing bullying | School-based | ICCS, PASEC, PIRLS, PISA, SACMEQ, TERCE, TIMSS | Student, principal, teacher, school |
| | | Household-based | Young Lives | Individual |
| Target 4.c. | By 2030, substantially increase the supply of qualified teachers, including through international cooperation for teacher training in developing countries, especially least developed countries and small island developing States | | | |
| 4.c.1 | Proportion of teachers in: (a) pre-primary education; (b) primary education; (c) lower secondary education; and (d) upper secondary education who have received at least the minimum organized teacher training (e.g., pedagogical training) pre-service or in-service required for teaching at the relevant level in a given country, by sex | School-based | EGMA/EGRA, ICCS, ICILS, PASEC, PIRLS, PISA, SACMEQ, TERCE, TIMSS | Teacher, school |
| | | Household-based | PAL Network | School |

| Indicator number | Indicator description | Type of assessment | Major learning assessments collecting needed data | Questionnaire |
|---|---|---|---|---|
| 4.c.2 | Pupil-trained teacher ratio by education level | School-based | EGRA/EGMA, ICCS, ICILS, PASEC, PIRLS, PISA, SACMEQ, TERCE, TIMMS | Teacher, school |
| | | Household-based | PAL Network | School |
| 4.c.3 | Percentage of teachers qualified according to national standards by education level and type of institution | School-based | EGMA/EGRA, ICCS, ICILS, PASEC, PIRLS, PISA, SACMEQ, TERCE, TIMMS | Teacher, school |
| 4.c.4 | Pupil-qualified teacher ratio by education level | School-based | EGMA/EGRA, ICCS, ICILS, PASEC, PIRLS, PISA, SACMEQ, TERCE, TIMMS | Teacher, school |
| 4.c.5 | Average teacher salary relative to other professions requiring a comparable level of qualification | School-based | PASEC | Teacher |
| 4.c.6 | Teacher attrition rate by education level | School-based | SACMEQ | School |
| 4.c.7 | Percentage of teachers who received in-service training in the last 12 months by type of training | School-based | EGMA/EGRA, ICILS, PASEC, PIRLS, PISA, SACMEQ, TERCE, TIMSS | Teacher |

*Source: UNESCO Institute for Statistics (UIS).*

The world is facing a crisis of learning, with many children leaving school without the basic skills they need for a prosperous and productive adult life. Two-thirds of the estimated 617 million children and adolescents who cannot read a simple sentence or manage a basic mathematics calculation are in the classroom. Too many are waiting for a quality education that never comes.

As the 2018 SDG 4 Data Digest shows, it is not enough to hope that they will stay in school and somehow acquire skills in reading and mathematics. It is critical to monitor those skills as children progress through school. That requires comparable data, over time, to ensure that children – and the education systems that serve them – are on track.

Given the critical importance of learning for the achievement of all the Sustainable Development Goals (SDGs), from poverty reduction to peaceful societies, this year's edition of the SDG 4 Data Digest is dedicated to the theme of learning outcomes. It showcases the most comprehensive and up-to-date compilation of work to inform the learning indicators of SDG 4.

The Digest discusses learning evidence on early child development, mathematics and reading skills among school-aged children, and digital and work-related skills among youth and adults. It highlights the conceptual frameworks and tools developed by leading authors and institutions to understand, measure, monitor and support learning for all. It also considers the implications of reporting for SDG 4.

United Nations
Educational, Scientific and
Cultural Organization

UNESCO
INSTITUTE
FOR
STATISTICS

SUSTAINABLE
DEVELOPMENT
GOALS

ISBN 978-92-9189-230-3

UNESCO Institute for Statistics
P.O. Box 6128, Succursale Centre-Ville
Montreal, Quebec H3C 3J7
Canada