# The role of big data in education quality monitoring: implications at the global, regional and national levels

Ursula Schwantner, Amit Kaushik, Ross Turner, Charlotte Waters & Ray Adams

***Abstract***

The paper discusses the role of big data in monitoring education quality and implications at the global, regional and national levels in the light of international efforts to monitor progress in achieving UN Sustainable Development Goal in Education (SDG 4). Data on learning outcomes are central in establishing and monitoring education quality. The proportion of children achieving at least a minimum proficiency in reading and mathematics is a primary indicator of SDG 4. For this indicator to be meaningful across contexts, a shared understanding must be reached on its constituents and the data used to report progress. While large-scale assessments are widely recognized as a primary source for such data, they vary in method and scope, posing major challenges for global monitoring. Consequently approaches to link major cross-national assessments and to harmonise quantitative data across such programs seem promising, despite their limitations in reach and in providing substantive information to inform improvements. In contrast, common described scales provide a reference point for data from a range of different assessments, be they international or national in scope, and including learning data on out-of-school children. The empirically derived scales describe learning progressions in reading and mathematics demonstrated by young learners. Building a bridge between statistical and conceptual approaches, the described scales enable big data from multiple sources to be used, and to be translated into meaningful descriptions of learning and targeted interventions. The scale development program also strongly focuses on capacity building. Observed challenges are articulated by the example of countries in South Asia that are still struggling to build assessment capacity. For global education monitoring to be effective, 'big data' need to be of high-quality, internationally consistent, and relevant for the education systems concerned.

## 1. Global education monitoring: quality versus quantity in using big data

Data on learning outcomes are central in establishing and monitoring education quality. For education systems, large-scale assessments are a well-established method to gain evidence on student learning to inform policy and practice (Masters, 2017; Best et al., 2013; Braun, Kanjee, Bettiner & Kremer, 2006). The monitoring of education quality is also central to the United Nation's Sustainable Development Goal in Education (SDG 4): *"Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all"*. The proportion of children achieving at least a minimum proficiency in reading and mathematics is a primary indicator of SDG 4 (Inter-Agency and Expert Group on Sustainable Development Goal Indicators, 2017).

For this indicator to provide meaningful information about the relative quality of the diverse education systems the major constituents (e.g., "minimum proficiency",

"reading" and "mathematics") need to be clearly defined to gain a consistent meaning across education systems. The same is true for the data used to report progress on a global scale (ACER-GEM, 2018; Jackson, Adams & Bernard, 2017).

International and regional assessments[1] play an important role in addressing these definitional challenges (Jackson, Adams & Bernard, 2017). Most importantly through the development of assessment frameworks that provide detailed descriptions of the learning domains and the constructs measured (for example OECD, 2015a; TIMSS 2015 Assessment Frameworks, 2013; Mullis & Martin, 2013). Based on rigorous scientific methods, assessment theory and practice (documented for example in OECD, 2015b; OECD, 2017; Olson, Martin & Mullis, Eds., 2008), international and regional programs are expected to provide data of adequate quality and are therefore widely recognized as a major data source for global monitoring. However, these programs do vary in content, method and scope so that comparability remains limited (Cresswell, Schwantner, Waters, 2015, Turner et al., 2018). Global monitoring based on major international and regional assessments would therefore only relate to groups of education systems participating in and using data from the same program for reporting.

Another major challenge in relying on international and regional programs in global monitoring lies in the reach of these programs. As an example, between 2010 and 2017 out of the 65 partner developing countries of the GPE[2], 54 (83%) participated in at least one kind of assessment of national scope (see Exhibit 1). If only data from regional or international programs was used for SDG 4 reporting, only 34 countries would be covered (63%), out of the 54 countries for which data on learning outcomes would be available. If national assessments were taken into account, 70% of countries would be covered. In addition, data from Early Grade Reading (EGRA)/Early Grade Mathematics

[1] Including the Programme for International Student Assessment (PISA) by the Organisation for Economic Co-operation and Development (OECD); Trends in International Mathematics and Science Study (TIMSS), the Progress in International Reading Literacy Study (PIRLS) by the International Association for the Evaluation of Achievement (IEA); the Programme d'Analyse des Systèmes Educatifs de la CONFEMEN (Conference des Ministres de l'Éducation des États et Gouvernements de la Francophonie) (PASEC); the South and Eastern African Consortium for Monitoring Educational Quality (SACMEQ); the Laboratorio Latinoamericano para la Evaluación de la Calidad de la Educación (LLECE) by the UNESCO Regional Bureau for Education in Latin America and the Caribbean; the Pacific Island Literacy and Numeracy Assessment (PILNA) by the Educational Quality and Assessment Programme (EQAP) of the Pacific Community (SPC); and the Southeast Asia Primary Learning Metrics (SEA-PLM) by the Southeast Asian Ministers of Education Organization (SEAMEO) and UNICEF East Asia and Pacific Regional Office (EAPRO).

[2] The number of GPE's developing country partners is 65 as of July 2017. Almost all of these are low-income countries, including countries that are affected by fragility and conflict. The number of children in the partner countries is estimated to be 330 million, of which 79 million are out of school (GPE, 2018b).

Assessments (EGMA)[3] and citizen-led assessments[4] could be used to fill gaps in reporting.

**Exhibit 1: Assessment programs in partner developing countries of the GPE in the period 2010–2017 (source: GPE, 2018a)**

| Total | None | At least one kind of assessment | National assessment | Regional assessment | International assessment | Citizen-led assessment | EGRA or EGMA |
|---|---|---|---|---|---|---|---|
| 65 | 11 | **54** | 38 | 26 | 8 | 9 | 40 |

A more flexible approach to global education monitoring should therefore consider data from a variety of assessments, including programs that provide information on children who are not attending school. Despite increased efforts to ensure access to education, the number of out-of-school children remains high in developing countries. In 2015, 264 million primary and secondary age children were out of school[5]. Given the stagnation in out-of-school rates and overall low completion rates[6] (UNESCO, 2017) it seems sensible that global monitoring takes into account this part of the population.

Another important aspect of global monitoring is to provide substantive information for education systems to inform improvements. Approaches that focus on the harmonisation of quantitative data across education systems often lack essential qualitative information and detail about what the numbers mean. A recent paper from Altinok, Angrist & Patrinos (2018) describes an approach where data from 163 countries and regions, covering developed and developing countries that participated in international and regional assessments over the period of 1965 to 2015[7], were compiled into a "global dataset on education quality". Multiple methods were applied to link achievement scores across programs and over time[8]. The resulting database consists of

[3] Research Triangle Institute (RTI) International

[4] Citizen-led, household-based assessments of basic reading and numeracy skills are currently undertaken in 14 countries in East and West Africa, South Asia and North America (Mexico) (http://palnetwork.org/)

[5] 61 million children of primary school age (9% of the age group), 62 million adolescents of lower secondary school age (16%), and 141 million youth of upper secondary school age (37%) (UNESCO, 2017, p. 118).

[6] In 2010–2015, completion rates were 83% for primary, 69% for lower secondary and 45% for upper secondary education (UNESCO, 2017, p. 118).

[7] International standardised assessments included are IEA/TIMSS and PIRLS and their precursors, the Monitoring Learning Achievement (MLA) program by UNESCO and UNICEF, and OECD/PISA. Regional standardised assessments included are SACMEQ, CONFEMEN/PASEC and UNESCO/LLECE. Hence the resulting database represents approximately two-thirds of developing economies (Altinok, Angrist & Patrinos, 2018).

[8] To link learning outcomes across programs and to include countries that only participated in a regional assessment, "doubloon countries" were used that participated in both international and regional assessments. In addition, international assessments conducted before the 1990s were anchored across time using the National Assessment of Educational Progress (NAEP) in the United States, which was conducted consistently from the 1990s onwards, and the U.S. participated in NAEP

rigorous and globally comparable estimates in reading, mathematics and science, including mean scores and achievement distributions in the form of low, intermediate and advanced proficiency thresholds for primary education (based on IEA/PIRLS and TIMSS) and secondary education (based on OECD/PISA). The main outputs that can be derived from the database are quantitative comparisons of results between education systems. The focus is thereby on achievement scores and achievement distributions that can be disaggregated across important equity indicators such as gender, socioeconomic status, language, urban/rural location and immigration status (Altinok, Angrist & Patrinos, 2018). What is missing however, is substantive information to inform the interpretation of the observed outcomes. For instance information about the consistency of the constructs measured in the three domains, or about the meaning of the observed differences in achievement. While substantive descriptions are provided for the proficiency thresholds, these are based on international assessments. As the authors mention, these might not be relevant or meaningful for developing countries, where less than 50% of students reach the minimum threshold (Altinok, Angrist & Patrinos, 2018).

For global education monitoring to have an impact, reporting of learning outcomes must be meaningful and inform system improvement. Reporting mechanism should therefore take into account substantive descriptions of learners' proficiency in the relevant domains (Adams, Jackson & Turner, 2017).

## 2. The concept of a Learning Progression

An approach to reporting assessment outcomes is needed that gives substantive meaning to the results for as many countries as possible. The ideal approach will also recognise that individuals develop at different rates within different subject areas and in comparison to their peers (e.g., Masters, 2018). As a result, the most useful approach will show progress, rather than only performance against a single benchmark.

A *Learning Progression* provides a good basis to meet this need, since it can support all learners to make good progress, and can be applied across all forms and levels of educational assessment (Waters, 2018). At the heart of this approach is the principle that assessment involves "the process of establishing where students are in their long-term learning and what progress they are making over time, usually in terms of their developing knowledge, skills and understandings." (Masters, 2013, p. 6).

Learning progressions conceptualise and describe learning domains with reference to core concepts that are present in basic forms in early learners, and become increasingly abstract, generative or sophisticated as learning progresses. They have a *horizontal structure* that identifies different aspects of the learning domain, and a *vertical structure* that describes and illustrates what learning progress looks like for the domain.

Learning Progressions are developed from empirical evidence of typical sequences of learning progress, and also from theoretical understandings of the nature of progress

as well as in various international assessments for the respective intervals (Altinok, Angrist & Patrinos, 2018).

(e.g., logical prerequisites for learning). They also are influenced by curriculum conventions and intentions (e.g., sequences in which material is commonly introduced), and informed by the experiences and practices of teachers.

By describing learning in substantive terms, Learning Progressions encourage understandings of learning levels and learner progress that numerical scores alone cannot convey. Learning Progressions encourage understandings of learning attainment and learning progress that are aligned with the objective of using assessment to support all learners, regardless of their starting points, to make good progress in their learning. This makes Learning Progressions particularly useful in developing countries, and in situations where high proportions of students are at relatively low levels of attainment.

Learning Progressions of this kind (Turner, Adams, Schwantner, Cloney, Scoular, et al., 2018) have been offered to the international community as a possible way forward in solving the global education monitoring challenge. A set of draft Learning Progressions was developed, for reading and mathematics, spanning learning that typically occurs from the earliest stages of schooling through to the later stages of secondary schooling and beyond. Turner et al. (2018) provide a detailed description of the process followed to develop those Learning Progressions. Some of the key elements are summarised here:

- A large number of test items were sourced from a wide range of mathematics and reading assessments in different countries[9], and used in a pairwise comparison study to generate item difficulty estimates on a single scale.
- Item demand for these items was analysed and used along with item difficulty estimates to draft descriptions of increasing levels of mathematics and reading knowledge needed by individuals undertaking these test items.
- Additional information from existing longitudinal assessment programs[10] was referenced to refine and expand the knowledge descriptions for reading and mathematics.
- Draft descriptions of increasing levels of knowledge for reading and mathematics were further refined with information from a wide consultation process of mathematics and reading experts.

The combination of qualitative and quantitative methods used to develop the Learning Progressions means that they are grounded in theory and supported by empirical data to ensure the breadth of the constructs, and their applicability across a wide range of learning contexts.

[9] Source assessment programs include ASER India (ASER Centre), EGMA and EGRA (RTI International), Australia's Longitudinal Literacy and Numeracy Study, Afghanistan's Monitoring Trends in Educational Growth program, the Online Literacy Assessment of Youth (Northern Territory, Australia), PILNA, IEA/PIRLS and TIMSS, OECD/PISA, SACMEQ, the Solomon Islands Standardised Tests of Achievement, and UWEZO (Twaweza) in Kenya, Tanzania and Uganda.
[10] Including ACER's extensively validated Progressive Achievement Tests for mathematics and reading.

These Learning Progressions are intended to provide a stable reference frame against which outcomes of different assessments in different countries can be aligned. Further empirical work is underway and planned, to establish the relationship between the outcomes of existing assessments, and the Learning Progressions. In this way, countries would be supported in using their existing assessments to establish measures of progress against a common frame of reference and to give those measures substantive meaning.

An important option of this empirical work is to test and compare results against the global dataset on education quality developed by Altinok, Angrist & Patrinos (2018) (see section 1). Provided that there is sufficient complementarity of outcomes, substantive meaning can be provided for the achievement scores in aligning these with the Learning Progression scales. Furthermore, as Altinok, Angrist & Patrinos (2018) indicate, alternative thresholds can be developed to increase the interpretability of the global database.

## 3. Big data in education monitoring: ensuring quality through increased capacity

For global education monitoring to be effective, 'big data' need to be of high-quality. Efforts around monitoring SDG 4 are therefore accompanied by numerous measures to strengthen countries' capacity in learning assessment. These range from knowledge sharing in international[11] and regional networks[12] to specific funding programs[13]. Observed challenges for countries in building assessment capacity are addressed in the following by the example of countries in South Asia.

The focus on learning assessments in South Asia is a relatively recent phenomenon. Bhutan, Pakistan and Sri Lanka have specifically taken note of international large-scale assessments and their potential for supporting improvement in learning (Pakistan, 2009; 2017; Bhutan, 2014; Sri Lanka, 2013). While Pakistan has committed itself to participating in TIMSS in 2019, Bhutan opted for PISA for Development (PISA-D) in 2017.

In India and Pakistan, public debate around learning outcomes can be traced to the publications of the Annual Status of Education Report (ASER) (Pratham, 2006; ASER Pakistan, 2008). Citizen-led in both countries, ASER provides a snapshot of student learning, testing children in the household, covering children in- and out-of-school. The findings from these assessments and the resulting debate in the popular media have led

[11] E.g., the Global Alliance to Monitor Learning (GAML), hosted by the UNESCO Institute for Statistics (UIS) (http://gaml.uis.unesco.org/);
[12]E.g., the Network on Education Quality Monitoring in the Asia-Pacific (NEQMAP), TALENT in Sub-Saharan Africa, or the Network for African Learning Assessment (NALA).
[13]E.g., the GPE Education Sector Program Implementation Grants (GPE, 2018c), or the planned GPE Knowledge and Innovation Exchange (KIX). Both initiatives focus on strengthening learning assessment systems (GPE, 2018b).

to changes in policy, shifting focus from inputs to learning outcomes, as in India (Planning Commission, 2012).

Although regional learning assessments have developed in Latin America (LLECE), Africa (SACMEQ and PASEC), Southeast Asia (SEA-PLM), and the Pacific (PILNA), there is no comparable regional assessment in South Asia.

Barring the exceptions noted above then, countries in South Asia tend to rely more on national assessments to provide a diagnosis of their education systems (UNESCO, 2015). The focus is thereby on classroom-based, formative assessments. For instance, India has been refining the domestic National Achievement Surveys (NAS) and developed the Continuous Comprehensive Evaluation (CCE) in the classrooms to provide relevant information on student learning.

Whichever the form of assessment, the use and analysis of big data must be based on high quality data in order for resulting decisions to be valid and meaningful (Cai, Zhu, 2015). But poor capacity and weak educational systems make the task of running large-scale assessments consistently and analysing their outcomes challenging.

A review of the experience of middle-income countries that participated in PISA between 2000 and 2015 found that many of these countries faced technical challenges, covering both analytical and operational aspects (Lockheed, Prokic-Bruer, Shadrova, 2015). Weak systemic capacity can often create unintended hurdles to successful implementation of standardised learning assessments in South Asia too. Based on ACER's work in the region, we can broadly classify observed capacity challenges in South Asia into three categories – institutional, technical and operational (see Exhibit 2).

**Exhibit 2: Three categories of observed capacity challenges in South Asia**

| Institutional | Technical | Operational |
|---|---|---|
| Political compulsions | Assessment framework, item development, translation | Standardised field operations |
| Technical skills | Technical standards | Record keeping |
| Human resource availability | Sampling | Quality monitoring |
| Budgets and timelines | Quality control | |
| University capacity | Data analysis and reporting | |

*Institutional challenges*

Policy goals for learning assessment are set by the political establishment. *Political compulsions,* such as the need to demonstrate educational progress or the impact of policy, frequently dictate the nature of assessment, the agency selected to undertake it, and the degree to which its outcomes are reported publicly. In addition, the *lack of technical skills* in education bureaucracy can lead to incoherence between the learning assessments and policy goals. Coupled with these challenges is the issue of *human resource availability.* Adequate staff is frequently not available, and those available are commonly rotated to other assignments, thus losing the benefit of continuity and stability along with the time and resources invested in capacity building. This has been

noted as one of the most difficult challenges in undertaking national assessment (Poudel, 2016).

Related challenges arise through inadequate *budgets* and unrealistic *timelines*. For example, data from an assessment undertaken in one country in 2013 was only analysed and reported upon in 2015 due to paucity of adequate financial resources. In another instance, lack of budgetary resources prevented printing of standard scanning sheets for student responses; the manual entry process used instead resulted in data from only half the sampled provinces being usable. Similarly, governments often delay internal decision-making processes beyond the planned start dates of assessments, yet expect that final outcomes will be reported by the scheduled dates.

Finally, one of the most significant institutional challenges lies in *university capacity.* Despite criticism, teacher training curricula of most countries in the region remain rooted in traditional concepts, with assessment considered largely in the context of constructivist, classroom-based practice (Gupta, 2018). As a result, subjects like advanced statistics for educational planning, data analysis or psychometrics are almost never taught at university level, thus adding further to the shortage of trained personnel for these areas.

*Technical challenges*

Technical capacity is a key requirement for developing and implementing robust assessments (ACER & UIS, 2017). Several learning assessments in the region have had to rely extensively on funding from international development partners and external technical support to undertake some or all of the assessment, with governments choosing to limit their engagement to planning and reporting. The technical support is mainly related to building the skill sets needed for *assessment framework and item development,* developing *technical standards, sampling, data analysis and reporting. Ensuring adequate quality* becomes a challenge, given the unfamiliarity of the participants with the stringent processes involved at each stage.

*Operational challenges*

Standardised field operations are essential to ensure learners' achievement is independent from the assessment conditions (ACER & UIS, 2017). Given the vastly dispersed geographies of some countries in this region and the differing infrastructure available in schools, ensuring *common standardized assessment procedures* poses a challenge. *Poor record keeping* and weak *quality monitoring systems* compound the problem of determining whether the final data is reliable and valid. A common misconception for example appears to be that providing a field operations manual is a sufficient and adequate measure, without training staff in actually implementing the assessment in a standardised way. Equally, there is often the misplaced confidence that existing government mechanisms of data collection can be used for gathering assessment data without too much modification.

## 4. Conclusion

For global education monitoring to be effective, 'big data' need to be of high-quality, internationally consistent, and relevant for the education systems concerned. Given the diversity of contexts of education systems, a focus in global monitoring on data from major international and regional assessments will fail to reach a large number of the most vulnerable education systems targeted by the Sustainable Development Goals in the first place. A flexible solution is therefore needed that is inclusive of assessments and learners, including out-of-school children. Furthermore, global monitoring needs to be based on an approach that adds substantive meaning to the numbers reported in order to inform improvement.

Common described scales in the form of Learning Progressions provide a reference point for data from a range of different assessments, be they international or national in scope, and including programs on out-of-school children. Moreover that approach inserts meaning into assessment outcomes and provides a clear basis for next steps in teaching and learning and in system improvement. The empirically derived scales referred to in this paper describe learning progression in reading and mathematics. Building a bridge between statistical and conceptual approaches, the described scales enable big data from multiple sources to be used, and to be translated into meaningful descriptions of learning and targeted interventions.

The common described scales also provide meaningful information to inform assessment reform and improvement. As such they have the potential to inform the various capacity-building initiatives accompanying the efforts of SDG 4 monitoring, contributing to high quality in assessment and reporting. As addressed in the example of countries in South Asia, technical capacities for diagnostic assessment remain weak, as does the ability to apply findings from such assessments to policy and practice. Although there has been an increased emphasis on systematic learning assessment in the region in recent years, the associated systemic capabilities are still being built.

## 5. References

ACER-GEM (2018). *The UNESCO Institute for Statistics Reporting Scales. Concept Note.* Australian Council for Educational Research. https://www.acer.org/files/UIS_Reporting_Scales_concept_note.pdf

ACER-GEM (2017). *System-Wide Analysis of Assessment Practices (SWAAP) Concept Note.* Working Draft. Melbourne: Australian Council for Educational Research.

ACER-GEM & UIS (2017). Principles of Good Practice in Learning Assessment. Australian Council for Educational Research Centre for Global Education Monitoring and UNESCO Institute for Statistics. http://uis.unesco.org/sites/default/files/documents/principles-good-practice-learning-assessments-2017-en.pdf

Adams, R., Jackson, J., Turner, R. (2017). *Learning progressions as an inclusive solution to global education monitoring.* Melbourne: Australian Council for Educational Research Centre for Global Education Monitoring (ACER-GEM).

Altinok, N., Angrist, N. & Patrinos H.A. (2018). Global Data Set on Education Quality (1965-2015). Policy Research Working Paper; No. 8314. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/29281 License: CC BY 3.0 IGO.

ASER Pakistan (2010). *Annual Status of Education Report (Rural) 2008.* Lahore: South Asia Forum for Education Development.

Best, M., Knight, P., Lietz, P., Lockwood, C., Nugroho, D., & Tobin, M. (2013). *The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries: Final report.* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Braun, H., Kanjee, A., Bettiner, E., & Kremer, M. (2006). *Improving Education Through Assessment, Innovation, and Evaluation.* Cambridge, MA: American Academy of Arts and Sciences.

Cai, L. & Zhu, Y., (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*. 14:2.

Clarke, P. (2017). *Making use of assessments for creating stronger education systems and imoproving teaching and learning.* Background paper prepared for the 2017/8 Global Education Monitoring Report. ED/GEMR/MRT/2017/P1/6 http://unesdoc.unesco.org/images/0025/002595/259564e.pdf

Cresswell, J., U. Schwantner and C. Waters (2015). *A Review of International Large-Scale Assessments in Education: Assessing Component Skills and Collecting Contextual Data*, PISA, The World Bank,Washington, D.C./OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264248373-en

GPE (2018a). *Strengthening Learning Assessment Systems.* A GPE Knowledge and Innovation Exchange (KIX) Consultation Paper. Draft for consultation (July 2018). Global Partnerhip for Education(GPE)

GPE (2018b). Global Partnerhsip for Education. Results Report 2018. Washington DC: The Global Partnership for Education. https://www.globalpartnership.org/content/results-report-2018

GPE (2018c). Guidelines for Education Sector Program Implementation Grants – ESPIG. Washington DC: The Global Partnership for Education. https://www.globalpartnership.org/content/global-partnership-education-program-implementation-grant-guidelines

Gupta, Latika (2018). Discourse of teacher education in India. In Kumar, Krishna (Ed.), *Routledge Handbook of Education in India – Debates, Practices and Policies*. Abingdon, Oxon: Routledge.

Inter-Agency and Expert Group on Sustainable Development Goal Indicators. (2017). *Report of the Inter-Agency and Expert Group on Sustainable Development Goal Indicators*.

Jackson, J., Adams, R. & Bernard, J. (2017). *Measuring learning: from comparison to understanding.* Global Partnership for Education Blog. Published 24/04/2018 https://www.globalpartnership.org/blog/measuring-learning-comparison-understanding

Kaushik, Amit (2018). Measuring Learning Outcomes in India. *Seminar,* June, Issue 706.

Lockheed, M., T. Prokic-Bruer & A. Shadrova (2015). *The Experience of Middle-Income Countries Participating in PISA 2000-2015.* Paris: OECD Publishing.

Masters, G. (2018). *A different way to organise the school curriculum.* Teacher Magazine. Australian Council for Educational Research. https://www.teachermagazine.com.au/columnists/geoff-masters/a-different-way-to-organise-the-school-curriculum

Masters, G. (2017). The science of large-scale assessment. In P. Lietz, J. Cresswell, R. Adams, & K. Rust (Eds.), *Implementation of large-scale assessments in education.* New York: Wiley

Masters, G. (2013). *Reforming education assessment: Imperatives, principles and challenges.* (Australian Education Review No. 57). Melbourne: ACER.

Ministry of Education (2014). *Bhutan Education Blueprint 2014-2024.* Thimphu: Royal Government of Bhutan.

Ministry of Education (2009). *National Education Policy 2009.* Islamabad: Government of Pakistan.

Ministry of Education (2010). *National Education Policy 2010.* Dhaka: Government of the People's Republic of Bangladesh.

Ministry of Education (2013). *Education First.* Battaramulla: Government of Sri Lanka.

Ministry of Federal Education and Professional Training (2017). *National Education Policy 2017-25.* Islamabad: Government of Pakistan.

Ministry of Human Resource Development (1986). *National Policy on Education 1986 (as modified in 1992).* New Delhi: Government of India.

Mullis, I.V.S. & M.O. Martin (Eds.) (2013). PIRLS 2016 Assessment Framework. TIMSS and PIRLS International Study Center and IEA, Chestnut Hill, MA and Amsterdam.

OECD. (2017). PISA 2015 Technical Report. Paris: OECD

OECD. (2015a). PISA 2012 Assessment and Analytical Framework: Mathematics, reading, science, problem solving and financial literacy. OECD Publishing. Retrieved from: http://dx.doi.org/10.1787/9789264190511-en

OECD. (2015b). PISA 2015 Technical Standards. Paris: OECD.

Olson, J.F., Martin, M.O., & Mullis, I.V.S. (Eds.). (2008). *TIMSS 2007 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Poudel, L. N. (2016). Reviewing the Practice of National Assessment of Student Achievement in Nepal. *Nepalese Journal of Educational Assessment*, 1(1).

Pratham Resource Centre (2006). *Annual Status of Education Report (Rural) 2005.* Mumbai: Pratham Resource Centre.

Planning Commission (2012). *Twelfth Five Year Plan (2012-2017) - Faster, More Inclusive and Sustainable Growth.* New Delhi: Government of India.TIMSS 2015 Assessment Frameworks. Copyright © 2013 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Turner, R., Adams, R., Schwantner, U., Cloney, D., Scoular, C., Anderson, P., Daraganov, A., Jackson, J., Knowles, S., O'Connor, G., Munro-Smith, P., Zoumboulis, S. & Rogers, P. (2018). *Development of Reporting Scales for Reading and Mathematics: A report describing the process for building the UIS Reporting Scales.* Australian Council for Educational Research https://research.acer.edu.au/monitoring_learning/33

UNESCO. (2017). Global Education Monitoring Report 2017/8. Accountability in Education: Meeting our Commitments. Paris, UNESCO. http://unesdoc.unesco.org/images/0025/002593/259338e.pdf

UNESCO. (2015). Education for all 2000-2015: Achievements and Challenges. Paris, UNESCO. http://unesdoc.unesco.org/images/0023/002322/232205e.pdf

UNESCO Institute for Statistics. (2018). Quick Guide to Education Indicators for SDG 4. Montreal: UNESCO Institute for Statistics. http://uis.unesco.org/sites/default/files/documents/quick-guide-education-indicators-sdg4-2018-en.pdf

UNESCO Institute for Statistics. (2017). *Cross-national Assessments technical meeting: Summary shared at the Global Alliance to Monitor Learning, Fourth Meeting.* November 2017. http://uis.unesco.org/sites/default/files/documents/gaml4-cross-national-assessments-technical-meeting-summary.pdf

UNESCO Institute for Statistics. (2016). *Laying the foundation to measure Sustainable Development Goal 4*. Montreal: UNESCO Institute for Statistics. http://uis.unesco.org/sites/default/files/documents/laying-the-foundation-to-measure-sdg4-sustainable-development-data-digest-2016-en.pdf

Waters, C. (2018) *Learning Progressions in ACER's Work*. https://research.acer.edu.au/monitoring_learning/34/