Inclusive and Equitable Quality Education for All: Towards a Global Framework for Measuring Learning?

E-forum Synthesis Report

Executive Summary

Over 900 participants from around the world joined the new IIEP Learning Portal's first e-Forum, titled ***Inclusive and equitable quality education for all: Towards a global framework for measuring Learning?*** These registrants—34% of them from national agencies or ministries of education and 42% of them from sub-Saharan Africa, with every other continent also represented—brought diverse perspectives to bear on the question of what role learning assessments may play in the context of the new Education 2030 and Sustainable Development Goals.

Over the course of the e-Forum participants shared their experiences and opinions on some of the key questions surrounding international and national learning assessments today. These different types of assessments serve different purposes. While international assessments are most useful for evaluating the effects of diverse national policies, national assessments are able to provide finer-grained analysis tailored to specific national characteristics and priorities. Here are seven takeaway points that arose from this global discussion:

1. • There is significant disagreement over whether an international assessment framework could be applicable to all countries.

2. • There are differing opinions on which competencies should be assessed, in either a national or international context.

3. • There is uncertainty about the role that digital literacy and technology should play in assessments.

4. • Participants were concerned with how international emphasis on learning assessments might affect national curricula and teacher practices.

5. • Developing effective assessments requires input from the regional, national, and even local levels.

6. • The most appropriate unit of analysis for an assessment depends on the assessment's purposes.

7. • There is a need to build capacity to understand assessment approaches and accurately interpret and use assessment data.

Through the e-Forum, participants were invited to participate in a series of live interactive video presentations with experts from around the world, and they were also able to access a set of resources on international and national assessments, part of the IIEP Learning Portal's large library of materials on issues relevant to planning for improved learning outcomes. This inaugural e-Forum showed the great promise that the IIEP Learning Portal holds for bringing together education decision-makers and other stakeholders from around the world, to think and plan together how to achieve quality education and improved learning outcomes.

Table of Contents

List of Figures

Introduction

The **IIEP Learning Portal**, a new interactive platform created to help decision-makers worldwide plan for quality education and improved learning outcomes, hosted an e-Forum from 16-27 November 2015 to explore how learning outcomes are assessed in different regions and education systems worldwide. Bringing together policy briefs on learning issues, on strategic planning, and on monitoring learning outcomes, as well as more than 1,000 resources in a searchable database, the IIEP Learning Portal offers—at no cost to users—comprehensive, up-to-date and relevant information on learning issues, from primary through secondary education. A series of e-Forums is envisioned as part of the Portal, in order to stimulate discussion and collaboration on learning issues among stakeholders around the world.

The decision to focus the IIEP Learning Portal's first e-Forum on the topic of learning assessments arose out of a recognition that the measurement of learning is central to discussions on how to remedy the global learning crisis. The Global Monitoring Report *Education for All 2000 – 2015* concluded that despite positive progress, the world has not yet reached the EFA goals. While great strides have been made to increase school enrolment and gender parity, there are still 59 million children out of school globally and around 100 million children who do not complete primary education. Inequality in education has increased, with the poorest and most disadvantaged shouldering the heaviest burden. Overall, the poor quality of learning at primary level still has millions of children leaving school without basic skills.

The new Sustainable Development Goals (SDGs) aim to ensure inclusive and equitable quality education and promote life-long learning opportunities for all, and to ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes by the year 2030. Basic skills in literacy and numeracy for all are the highest priority. In order to monitor achievement of these goals, there is a need to discuss a strategic framework for international assessments. This can help to ensure reliable information on the efficiency of education systems and provide comparable indicators to show development over time.

Indeed, the most recent Global Monitoring Report shows that governments have increased efforts to measure learning outcomes through national and international assessments. The Learning Metrics Task Force, convened by UNESCO Institute for Statistics (UIS) and the Center for Universal Education (CUE) at Brookings, has discussed how to improve learning outcomes by strengthening assessment systems and using assessment data. The first phase of their work is presented in the report: *Toward Universal Learning – A Global Framework for Measuring Learning* (2013).

The IIEP Learning Portal e-Forum **'Inclusive and equitable quality education for all: Towards a global framework for measuring learning?'** built on previous initiatives in the field by looking at existing assessments of learning outcomes, and by encouraging discussion on the extent to which these assessments meet the needs for monitoring learning outcomes in different regions and education systems across the world. The two main themes for the e-Forum discussion were:

Theme 1: Large scale international assessments – one size fits all?

Large-scale international assessments are designed to gather information on learning outcomes in a way that is comparable across educational systems, and which can provide trend data for learning outcomes over time. The assessments consist of cognitive tests, accompanied by surveys of pupils and school administrators. Results are presented as comparable proficiency levels across countries.

Recently, both the Organisation for Economic Co-operation and Development (OECD) and the International Association for the Evaluation of Educational Achievement (IEA) have started to develop modified versions of tests like PISA, TIMSS and PIRLS for developing countries. Under this theme, e-Forum participants discussed to what extent existing large scale assessments can form a basis for monitoring the SDGs regarding the achievement of basic skills for children and youth in all countries, focusing on the following three questions:

1. How do you see the main advantages and challenges of existing international large-scale assessments?

2. How can modified versions of existing assessments (e.g PISA for Development, TIMSS Numeracy and PIRLS Literacy) be made relevant for countries?

3. What are the advantages of using regional approaches (e.g. SACMEQ, Pasec, LLECE) to assess learning outcomes?

Theme 2: National assessments for learning outcomes – how can they stimulate improved learning?

In the past two decades, national assessments have emerged as an important tool for providing a measure of educational achievement, and there are a great variety of national assessment programs, each with different aims and purposes. Broadly, one can differentiate between assessments that are designed for accountability at all levels, and assessments that are designed for system evaluation and development or designed to give feedback to parents and students. In both cases, the desire to measure change in achievement over time will imply specific requirements to the test design. Under this theme, e-Forum participants discussed how national, regional and international assessments can act together to form a robust framework for the monitoring of learning outcomes at all levels—pupils, schools, regions and countries. Discussions focused on the following three questions:

1. How can national assessments and exams be designed to give feedback to pupils, teachers and schools on the quality of learning?

2. How could international tests be used as a supplement to national assessment systems?

3. How to ensure that every country has the necessary competence and capacity to develop and implement a comprehensive framework for learning assessment?

Special guest presenters

During the two weeks of the e-Forum, participants were also invited to participate in a series of live interactive video presentations with experts from around the world. These talks featured Margarita Poggi (Director, Buenos Aires office of UNESCO International Institute for Educational Planning), Dr. Silvia Montoya (Director, UNESCO Institute for Statistics), Dr. Asmah Ahmad (President, Malaysian Association of Social Impact Assessment), Dr. Sarah Howie (Director, Centre for Evaluation and Assessment at the University of Pretoria, South Africa), and Dr. Sara Ruto (Director of the Secretariat, People's Action for Learning Network). Participants were able to watch live video streaming, submit questions via chat, and also access the recorded videos and power point presentations later on.

Where do we go from here?

In the concluding days of the e-Forum, all participants were invited to join a common discussion on a new set of questions that arose from the two main themes. The topics and questions addressed in the final days of the e-Forum were:

1. The "unit of analysis" for assessments: How detailed should the feedback from national and international assessments be? Put another way, what should the "unit of analysis" be—individual students, schools, sub-national regions, or nations?

1. Cost and capacity issues in implementing national and international assessments: What needs to be done in order for all countries to be able to conduct effective national assessments and participate in some kind of international assessment? How can these costs be met? How can it be ensured that each country has the necessary technical expertise and implementation capacity?

2. Use of assessment data, feedback mechanisms, monitoring systems: Ideally, how should international, national, sub-national, and specific educational institutions work together to use assessment data as part of an overall education monitoring system? What sorts of feedback pathways and mechanisms should exist in order to translate assessment data into improvements in the quality of education and learning outcomes?

Outline of this Synthesis Report

Over 900 participants from all over the world registered for this e-Forum. This synthesis report provides detailed information on the profile of these participants, the discussions regarding Theme 1 and Theme 2, what was learned from the special guest presentations, and the key points of agreement and disagreement that emerged from the discussions.

Participants

In total, 910 participants registered for the e-Forum. Each registrant filled out a brief questionnaire, giving some insight into the profile of those who were interested in participating.

Region: There was clearly global interest in the e-Forum, with the strongest presence from Africa (42%) and the Asia/Pacific (22%) regions. There were also registrants from Europe (15%), North America (9%), the Arab States (7%) and Latin America and the Caribbean (5%). The relatively lower participation from Latin America was likely due to the exclusive use of English in the e-Forum; future e-Forums are planned to take place in both Spanish and French to offer greater access to other parts of the world.

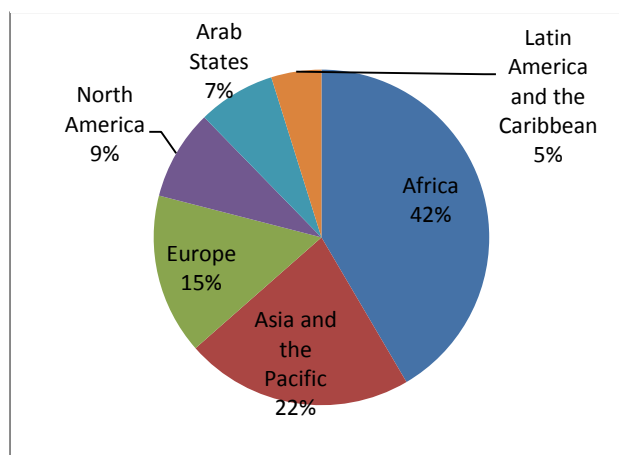**Figure 1: Geographical spread of e-Forum registrants (N=910)**

Figure 2: Map of e-Forum registrants' locations

Sex and Education: The registrants were 43% female, and the majority of all registrants (59%) held a Master's degree as their highest educational qualification. Another significant percentage of registrants held a Ph.D. (21%), while others had a Bachelor's degree (12%), a lower-level teaching qualification (3%) or some other type of education (5%).

Employer: The largest percentage of registrants reported that they currently work for a national agency or ministry of education (34%), with significant representation also from international organizations (16%), universities (15%), and NGOs (13%). A smaller number were employed at primary or secondary schools (4%) or in a non-education government ministry (2%).

Figure 3: Registrants' current employer by type of organisation
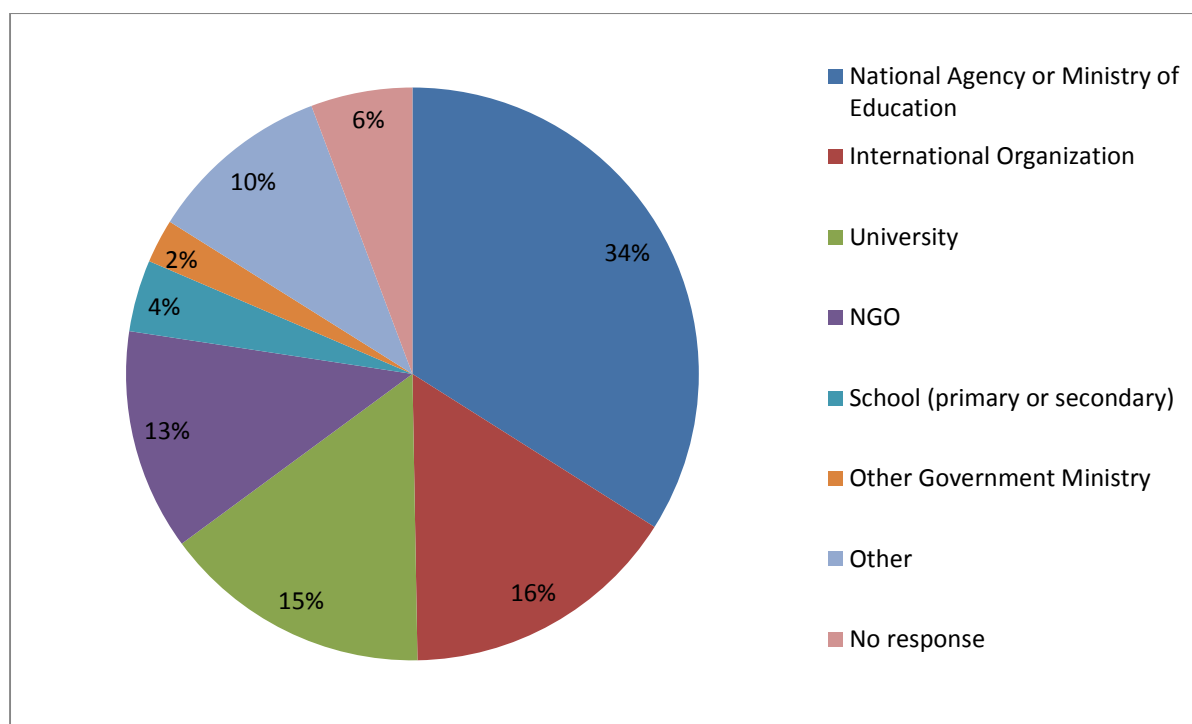
Registrants also described the professional title of their current position, although more than one quarter of participants selected "Other" or did not report their title. The largest percentage of registrants (37%, indicated below in dark blue) were in a managerial-level position, with 14% the Head of their respective office, division, section or unit; 10% Project Managers; 9% Department Directors; and 5% Assistant Directors. Other titles included Technical Advisor (8%), Researcher (8%), Professor (6%), Teacher (6%), Official (3%), Inspector (2%), Student (2%), Head Teacher (1%), and Research Assistant (1%).

Figure 4: Current title/position of registrants
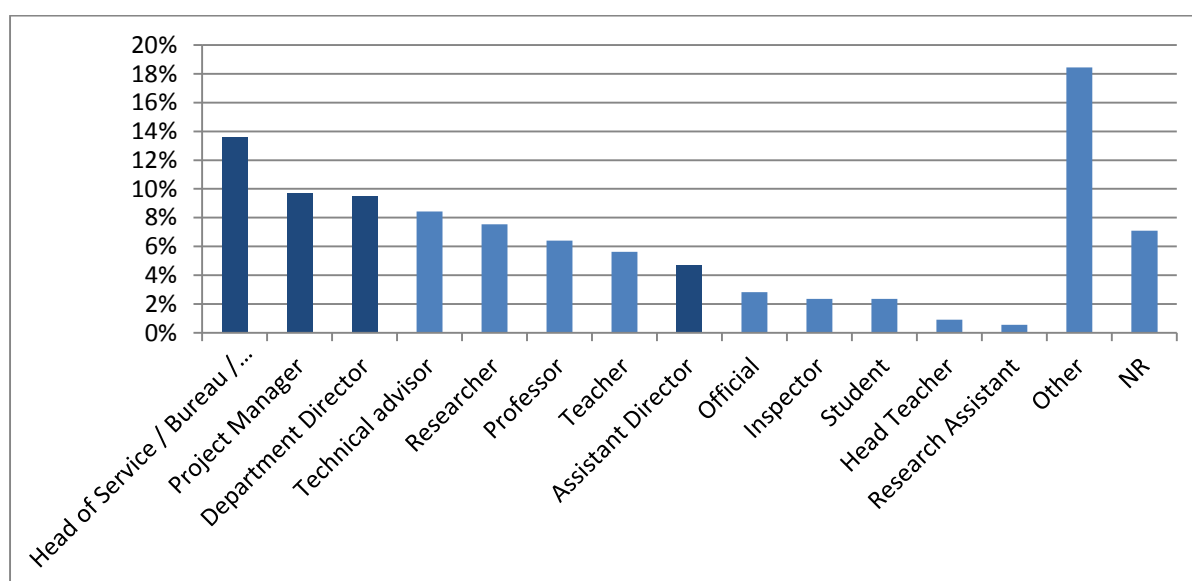
Out of the 910 registered participants, 356 (39%) logged in to use the e-Forum platform and its resources in some way. An average of 113 participants (32% of those who logged in) viewed each

special guest presentation either live or via the recording on YouTube, with the highest receiving 206 views and the lowest receiving 73 views. Finally, 70 participants (20% of those who logged in) participated actively in the discussion threads by posting at least one comment. It should be noted that there is some indication that others were also following the e-Forum via e-mail and Twitter. An evaluation survey was distributed to all registrants after the close of the e-Forum, in part to gather information regarding how to encourage a greater percentage of registrants to log on to the portal, join the live guest presentations, and actively participate in discussions.

Theme 1: Large scale international assessments—one size fits all?

Theme 1 of the e-Forum focused on large-scale international assessments such as PISA, TIMSS, and PIRLS, as well as the regional assessment programs that exist in several parts of the world. The discussion focused on three topics: (1) the main advantages and challenges of existing international large-scale assessments; (2) how modifications might make these assessments more relevant for particular countries; and (3) the advantages of using regional approaches to assessing learning outcomes. The following paragraphs summarize the outcome of the e-Forum discussions on each of these themes.

1. How do you see the main advantages and challenges of existing large-scale international assessments?

Participants identified two general advantages of large-scale international assessments: their potential to provide a global learning assessment framework that can help citizens hold governments accountable for providing a quality education, and the parallel between international assessments and the global competition students are facing in other domains, particularly employment.

According to some participants, a key advantage of international large scale assessments is that they help give a comparative picture of learning outcomes, which is important in the modern globalized world, and they can offer comparable, evidence-based and policy-relevant data that countries can rely on to help evaluate their education systems. International large scale assessments can be used to identify countries/regions where learning outcomes are poor, so that remedial action can be taken. A few participants commented that it is both possible and desirable that all countries should hold at least some educational objectives in common, since a global learning assessment framework could help citizens hold their governments accountable for providing a quality education. In order to do so, they said, the performance of education systems should be analysed down to the level of sub-national divisions. A participant from Nigeria argued that a local or regional approach should be the starting point for developing an assessment framework, but after that, global common assessment would be most desirable, since it can encourage regional collaboration.

Participants also considered to what extent existing large scale assessments can form a basis of tools to monitor the SDGs related to the achievement of basic skills for children and youth in all countries. It was argued that international assessments are key since we 'expect our children to compete internationally', and that the main advantage is that students participating in international assessment systems are exposed to types of questions and skills considered globally accepted, hence 'they are prepared [with] the basic knowledge and skills in order to excel their pursued career internationally' (Mariam Othman, Malaysia). Following further clarification of the e-Forum's definitions of learning outcomes and assessments, discussion began following two intertwined strands about whether it is possible, and desirable, to build a common, global learning assessment system for all.

However, a number of participants also pointed out disadvantages of the existing large-scale international assessments. In particular, they questioned whether developing countries have the expertise and resources to join international assessments, and also whether international tests take into account the specific challenges and situations in developing countries. Several participants argued against a global universal learning assessment system due to cultural differences in what and how humans learn. A major concern was variations in the learning outcomes and purpose of education, and it was argued that global assessments such as PISA, PIRLS, and TIMSS do not take national contexts

into consideration. The feasibility of achieving a universal assessment scale was questioned, especially with regards to implementation in developing countries.

As discussion developed, participants began to identify educational challenges within their own countries, considering the relevance, advantages and disadvantages of international and regional systems of assessments in these contexts. Many participants pointed to the vast differences between countries and education systems, particularly focussing on the differences between developed and developing countries, and what the implications are for curricula and therefore assessment. Despite recognising the potential for international assessment to cultivate higher educational standards within 'struggling' national systems, many questioned whether it is possible to define a common set of basic skills and competences for learners across the world. It was also pointed out that developing countries might not be prepared for these assessments because they may not have the required expertise and resources to implement them effectively. International tests must take into account context-specific challenges.

2. How can modified versions of existing assessments be made relevant for countries?

Participants took stock of existing experiences – both positive and negative – regarding international and national assessment systems in their own countries in order to consider how modified versions of existing assessments might be made relevant for countries. Acknowledging the huge challenges countries like Kenya and Nigeria are facing in order to build up quality education for all their children, Anne-Berit Kavli (Moderator) asked to what extent participants believe existing regional assessment programs (E.g. UWEZO, PASEC or SAQMEC) or international adaptations like TIMSS Numeracy, PIRLS Literacy at the primary level or PISA for Development at secondary level could provide relevant benchmarks and stimulate development in individual countries.

Many participants recognised the need to look at curricula in order to match this with suitable assessment, and through this topic philosophical discussions were had examining fundamental educational concepts, such as what the aims of education are, and whether they are the same internationally. Participants drew attention to the distinction between learning and assessing curricula based primarily within the developing world, such as maths and literacy skills, in contrast to more globally suited 21st century skills for the workplace such as digital literacy. Isatou Ndow (The Gambia), for example, thought that one could look at the state and needs of certain countries in order to modify assessment systems accordingly, although 'there are themes and topics that can be adapted easily while there are others that have cultural dimensions would need complete modification'.

The importance of considering teachers in the debate was re-iterated by a number of participants, in particular in relation to issues in developing countries with untrained teachers unable to design and evaluate assessment systems effectively. Some suggested that it would be beneficial to provide the teachers with in-service training courses and seminars in order to increase their skills in the area, and encourage them to adapt international assessments to national contexts and curricula. However, some participants referred to the huge challenges faced by many developing countries, and the need for some basic and acceptable common standards that can be valid across all different countries was underlined. Many again questioned whether all countries should have the same curricula due to contextual differences, yet at the same time a few participants argued for the necessity of a global curriculum to address global issues to create a sustainable society, such as climate change.

In the context of this discussion, some of the existing attempts at modification were considered, such as TIMSS Numeracy, PIRLS Literacy, and PISA for Development.

3. What are the advantages of using regional approaches to assess learning outcomes?

The majority of participants in the Theme 1 discussion directly addressed the problems inherent in designing an assessment system to be 'one size fits all', and pointed out the advantages of using regional approaches to assess learning outcomes instead. Many participants discussed the importance of considering the vastly different country-specific contexts when establishing assessment systems. Chidi Ezegwu (Nigeria) argued that regional assessment should replace global assessment since international systems do not accommodate local conditions, local manpower needs and socio-economic and study conditions in different countries.

There was also discussion of how international and national tests can supplement each other to build up a comprehensive monitoring system, with the proposal that existing regional assessment programs (E.g. UWEZO, PASEC or SACMEQ) or international adaptations like TIMSS Numeracy, PIRLS Literacy at the primary level or PISA for Development at secondary level can provide relevant benchmarks and also stimulate development of relevant national assessment systems in participants' own countries. Although some participants pointed out that country-specific assessments may not be consistent with the over-arching international assessment frameworks, many thought that a solution could be combining international assessment with modified versions for national use. Julius Bjornsson (Norway) raised his concerns about polarising comparative international big tests, and regional/national ones, since he argued that they should complement each other. He suggested that national systems should be based upon international ones, perhaps in terms of methodologies or topics tailor-made for individual countries' contexts and goals.

Technology in assessment

Marwa Biltagy (Egypt) brought up the advantages and disadvantages of incorporating technology into large-scale assessments. Advantages include reduced costs of data entry, collection and aggregation; the ability to adapt tests to students individually; and the ability to provide curriculum developers, researchers, teachers, and even students with detailed information that can be used to improve future learning. However, the main challenges include substantial costs for assessment systems, for example, devices and network, technical support and maintenance; and the need to integrate applications and systems so that standardized information can be collected and aggregated.

Theme 2: National assessments—how can they stimulate improved learning?

The Theme 2 discussion addressed three major topics: (1) how national assessments can be designed to give feedback on the quality of learning; (2) how international assessments could be used as a supplement to national assessment systems, as well as to national examinations and in-class formative assessment; and (3) the necessary competencies and capacities to develop and implement a national learning assessment framework. One additional topic—the question of what competencies should be measured on national and international assessments—generated a great deal of discussion.

1.   How can national assessments and exams be designed to give feedback to pupils, teachers and schools on the quality of learning?

As part of this discussion, participants gave details on several national assessment and examination systems, including in Iran, Ivory Coast, Kenya, Namibia, Rwanda, Singapore, Vietnam, and Zambia. Participants clearly agreed that national assessments should be designed to give feedback—particularly to pupils, teachers, and education decision-makers—in order to improve the quality of teaching and learning. Brigitte Matchinda (Cameroon) wrote, "In order to give feedback to pupils, teachers, and schools on the quality of learning, national assessments and exams should be designed to assess major aspects of the educational system functioning with respect to access, quality, efficiency and equity. The instruments should be properly designed to capture knowledge, skills, attitudes and habits that students have acquired on the one hand and on the other, the knowledge, pedagogical skills and attitudes of teachers."

Whether this feedback system works in practice, however, was a more difficult question. In some countries, it was pointed out, there are as yet no national assessments that are specifically designed to provide comparable results from place to place or year to year. In other countries, standardized national assessments do exist, but the way the resulting data is used does not seem to lead to clear improvement over time. Jenna Watson (UK) remarked: "Of the three contexts that I've lived and taught in I can't think of a case where this [feedback] operates on a systemic level. […] There still seems to be difficulty in using the data to really improve teaching and learning."

Three particular solutions were proposed to this problem. The first was the observation that teachers and other educators may need to be trained in order to effectively interpret and use assessment data to improve their practices. A second proposal that received extensive discussion was the idea that national assessments would be more successful in providing usable feedback if they were created through a bottom-up process of design in the first place, in which students, community members, teachers, and schools were involved in deciding what should be tested and how. Many participants agreed that involving more stakeholders in decision-making would help to ensure that the resulting data are relevant to the particular needs of each country or community. Third, some participants proposed that in order for assessment data to feed back into quality improvement, more attention needs to be given to how findings are communicated to different audiences.

2.   How could international tests be used as a supplement to national assessment systems? Also: what is the relationship to national examinations, other summative assessments, and in-class formative assessment?

During the discussions, it became clear that many participants see a significant overlap among international assessments, national assessments, national examinations, other types of summative assessments, and even in-class formative assessment. In some cases, the actual distinctions between

each of these approaches may have been unclear to participants. In other cases, however, participants were aware of the differences but were specifically arguing that there should be greater coherence and complementarity among them.

Symon Chiziwa (Malawi) argued that "international tests can simply give the general picture of how the education system is faring as compared to other countries," although countries can use them to learn about how to create higher-quality national examinations to meet their own needs. Denise Edwards (Commonwealth of Dominica) clearly stated that both international and national assessments have their uses. "As citizens of the global world," she wrote, "it is imperative that all learners develop some core competencies which will first be useful in building his/her community then the wider society. However, we must be cognizant of culturally sensitive factors which would require differences in assessment." She further mentioned that although "identifying core competencies and ways of assessing them based on similar standards would [lead to improvement] in many developing countries," many countries "would not encourage [participating in international assessments] because of not wanting to get out of the comfort zone." Participating in international assessments may be controversial for countries with poorer quality education systems precisely because they call attention to just how much needs to be improved.

Many participants seemed to share the view that, in order to improve learning outcomes, students need to receive direct feedback on their performance, akin to formative assessment. Yet as Tom Loveless (Moderator) pointed out, national assessments "usually are not timely enough to serve as formative assessments [since] results get back to schools several weeks or months after the assessment." Reflecting on this limitation, some participants argued that in-class formative assessment actually plays a more crucial role than national assessments. Deepa Idnani (India) argued that teachers need better preparation in order to conduct such formative assessments within the classroom, providing teachers with "certain guidelines, training, or resources" in order to help them understand "how to assess these core competencies."

Yet there are also some drawbacks to asking teachers to become too focused on assessment, whether at the international, national, or school levels. Wei Shin Leong (Singapore) explained that this can result in "a certain level of fragmentation and unauthenticity in teaching and learning", arguing that "education experts and scholars in psychometrics and formative assessment need to work together to devise ways and means to support schools and teachers in developing a balanced assessment system."

3. What are the necessary competencies and capacities to develop and implement a comprehensive framework for learning assessments? How can a country be sure to get them?

Discussions also touched on the question of how countries can develop the necessary capacity to implement an effective national learning assessment framework. Abdoul Carim (Niger) cited an OECD publication, *Synergies for Better Learning: An International Perspective on Evaluation and Assessment* (2013), which offers a clear description of the capacities that countries need in order to effectively design, implement, and make use of assessments, including the key importance of "building a coherent and integrated evaluation and assessment framework." Jason Pennells (UK) further listed the importance of "advanced psychometric skills, as well as adequate IT infrastructure and skills, wider assessment expertise and organizational management capacity at central and distributed levels."

There was general acknowledgment that many countries do not have deep expertise available in these issues.

Several participants emphasized the impact of information and communications technology (ICT), both on the global availability of information and thus on general learning conditions, as well as specifically on assessment in the form of computer-based tests. Olalekan Saidi (Nigeria) wrote, "Most developing countries still faced challenges of infrastructure deficit, lack of capacity on the part of both teachers and student to use computer, political will to drive the process etc. […] For instance, when [the government] introduced e-examination assessment to overcome malpractices in Nigeria, it created a heated debate among the states of the federation which are at different levels of development." Maiga Seydou (Ivory Coast) followed up on the discussion of assessment technologies by reminding the participants, "Whatever the method used, make sure that you produce reliable results."

Symon Chiziwa (Malawi) argued that joining an international assessment is one way for a country to build its own capacities for national assessment. If national assessments "begin to test some of the skills that international tests focus on," he wrote, "international tests can help enhance the quality of national tests." Symon and other participants remarked that a better understanding of assessments needs to be built at the university level, so that countries have a larger number of people who are trained in the specialties of test design and data interpretation.

Are there any skills and competencies that all students around the world should be expected to develop? Are there skills and competencies that are not held in common in different national contexts and therefore difficult to measure internationally?

A final fourth topic generated a great deal of interest among participants: the question of which skills and competencies should be assessed at the national or international levels. While there was general agreement that literacy and numeracy continue to be core important areas of learning, the vast majority of participants did not think that assessments should be limited to these areas. Marwa Biltagy (Egypt) wrote: "The objectives of education can no longer simply be to provide basic literacy skills for the majority of students. The skills that are easiest to teach and easiest to test are now also the skills that are easiest to automate, and outsource." She and other participants proposed a range of alternative or additional abilities that should be the focus of assessment, including problem-solving and creativity, inter- and intra-personal skills, citizenship, collaboration, vocational and employment skills.

Despite the high degree of importance participants ascribed to these abilities, they were unsure of whether international or national assessments could effectively measure them. The discussion moderator, Dr. Tom Loveless, remarked "Although I agree that critical thinking, problem-solving and creativity are important, they are extraordinarily difficult to assess. […] Most schools have a difficult time teaching all students how to read and some basic mathematics. After 100 years of trial and error, we have fairly good (but not perfect) tools for measuring those skills and knowledge. Maybe we should get that right first."

Special Guest Presentations (Hangouts)

The e-Forum was enriched through the contributions of five special guest presenters, speaking on topics related to national, regional, citizen-led, and international assessments: Margarita Poggi

(Director, Buenos Aires office of UNESCO International Institute for Educational Planning), Dr. Silvia Montoya (Director, UNESCO Institute for Statistics), Dr. Asmah Ahmad (President, Malaysian Association of Social Impact Assessment), Dr. Sarah Howie (Director, Centre for Evaluation and Assessment at the University of Pretoria, South Africa), and Dr. Sara Ruto (Director of the Secretariat, People's Action for Learning Network). The following is a summary of each presenter's main points.

Evaluation of Educational Quality: National Systems and International Studies in Latin American Countries

Margarita Poggi, Director of the Buenos Aires office of IIEP/UNESCO

Margarita Poggi discussed the effects of Latin American countries' widespread participation in regional and international studies to evaluate educational quality since the late 1980s and 1990s. This has led to some conceptual changes in educational evaluation in the region, such as understanding that evaluation involves value judgments and requires a multidisciplinary approach. There have also been methodological changes, in terms of the acceptance of the need to combine quantitative and qualitative methods, and debates on the nature of causality. Important points that are still being discussed include which areas of student achievement to evaluate, whether to evaluate school performance directly or indirectly, which grade(s) or year(s) to evaluate, the coverage or sample to be used for an assessment, and the frequency.

International studies offer good potential, Ms. Poggi explained, in creating longitudinal comparisons and synchronic comparisons between countries. However, they are limited by their macroscopic perspective and they need to be contextualized either nationally or regionally. Additionally, the meaning of assessment data can only be interpreted in connection with specific knowledge of the characteristics of each system. Some of the challenges that remain include: reviewing the design and strategy of evaluation from a rights perspective; expanding the range of abilities to be evaluated; reviewing the balance between high-stakes evaluation and formative evaluation; improving coordination between standardized external evaluations and internal evaluations; better integrating different sources of information on educational quality; and producing ad hoc systems for monitoring and evaluating educational plans and programmes.

The UIS Assessment Database and Catalogue: The First Step Towards a Global Partnership for Learning

Dr. Silvia Montoya, Director of the UNESCO Institute for Statistics

Dr. Silvia Montoya discussed how to monitor the progress of the new Sustainable Development Goals, and addressed the question of developing some basic, common standards for literacy and numeracy. She underlined the right to quality education for all children, and pointed to the need for measurement tools that focus on well-defined outcomes and related benchmarks, and on the measuring of trends and development. The Sustainable Development Goals are complex, and while at the global level there may be just one or two indicators to be measured for each specific sub-point under the Education goal (Goal 4), there are up to 43 thematic indicators that countries can make use of for more detailed measurements.

Dr. Montoya then presented the UIS effort to catalogue the national assessments that have been implemented around the world since the year 2000, and explore how these assessments might tie into the global effort to monitor the Sustainable Development Goals. The majority of countries now employ at least one type of national assessment, though there is heterogeneity in the use of these

assessments as planning and monitoring tools, and there is some question about the technical rigor of some assessment approaches. In order to build a global system for monitoring learning, the best path may be to define benchmarks and standards based on nationally-defined standards, and then develop a global mechanism to measure and frame these indicators at one point in time. Eventually, the goal might be to create a Universal Learning Scale, with initial field trials and data collection from 2021-2026. In the end, it is important to focus most on what the users of assessment data need in order to better understand their own education system processes and evaluate their impact.

Southeast Asia Primary Learning Metrics (SEA-PLM)
Dr. Asmah Ahmad, President of the Malaysian Association of Social Impact Assessment

Dr. Asmah Ahmad described the evolution of the newest regional assessment programme, the Southeast Asia Primary Learning Metrics (SEA-PLM). This assessment, she explained, is intended to improve and redefine learning outcomes in a culturally-appropriate and regionally-relevant manner. It is intended to enhance the capacity to generate assessment data at various levels, and to utilize assessment data for education quality improvement. The initial focus of the assessment is in the areas of literacy, numeracy, and global citizenship at the level of Primary Grade 5.

Dr. Ahmad pointed out that some unique features of SEA-PLM include its direct relevance to the curricula of the Southeast Asian region, the fact that it assesses writing across a number of different languages, and that it includes a focus on global citizenship and values characteristic of the region. It is planned that in 2016-2017 at least six countries from the region will pilot the assessment tools, and by 2020 all SEAMEO and ASEAN member countries will join. In order for this effort to succeed, Dr. Ahmad concluded, there is need for strong political leadership and commitment, adequate and predictable funding, alignment with government priorities and planning cycles, a coherent coordination mechanism, capacity development and technical leadership to ensure that the assessment meets international standards, a comprehensive communication strategy, and a commitment to use the resulting data in order to improve learning outcomes over time.

Experiences of the IEA's International Large-Scale Assessments in South Africa
Dr. Sarah Howie, Director of the Centre for Evaluation and Assessment, University of Pretoria

Dr. Sarah Howie presented on the South African experience participating in the international IEA studies PIRLS, SITES, and TIMSS. In the context of South Africa's adoption of the Education for All Framework for Action, there was a national commitment to improving educational quality and ensuring that measurable learning outcomes were achieved by all (EFA Goal 6). However, South Africa faced the challenge of heterogeneity in its education system with schools serving students from a diverse range of languages and cultures, and with many high-poverty schools—lacking even the minimum necessary conditions for learning—alongside very well-resourced schools. One major question for South Africa, therefore, was "How do we ensure that our measures are valid and reliable given the diversity of environments, cultures, economies, and policies?"

Dr. Howie described how, following the surprisingly low performance of South African Grade 8 students on TIMSS in 1995, 1999 and 2003, the government skipped the 2007 round of testing and then re-joined in 2011, but this time assessing Grade 9 instead of the international target Grade 8. Similarly, after performance on PIRLS was lower than expected in 2006, even while assessing in South Africa's 11 official languages, South Africa decided to instead employ the PIRLS Literacy (pre-PIRLS) test, in the hopes that it would be more suited to students' level of reading achievement. Despite

these adjustments, however, South African students have continued to perform poorly. Dr. Howie demonstrated, using excerpts from the media, how these experiences in international assessments have led to increased public awareness of the nation's education quality challenges.

PAL Network: People's Action for Learning

Dr. Sara Ruto, Director of the Secretariat for the People's Action for Learning Network

Dr. Sara Ruto presented on the origins and experiences of citizen-led assessments, from the first experience in India to today's global People's Action for Learning network. The first citizen-led learning assessment was conducted by India's largest NGO, ASER, in 2005. The approach was household based, rather than school-based, allowing assessments to reach even out-of-school children aged 7-14. The results showed a shockingly poor overall level of learning in literacy and numeracy. Although there was initial government denial, the consistency of findings in subsequent assessments has now stimulated serious response from national and state authorities.

Dr. Ruto described how coverage of the ASER experience in India led to similar initiatives in Pakistan, then Uganda, Kenya, and Tanzania, then Mali, Senegal, Mexico and Nigeria over the course of the past eight years. To support the growth, expansion, and visibility of these citizen-led assessments, the People's Action for Learning Network was founded in January 2015. The PAL Network's current goal is to grow a network of 25 countries conducting citizen-led assessments. The benefits of this approach, Dr. Ruto explained, include: reaching all children (not just those who are enrolled in school); involving parents so that they become more aware of education quality issues and more likely to advocate for their children's learning; the ability to collect evidence at sub-national levels, allowing for finer-grained interpretation of the results; and the commitment to conduct annual assessments in order to drive change more rapidly. Dr. Ruto also gave detailed information on how the PAL Network seeks to communicate its findings in order to be as accessible as possible to the public, with the ultimate goal of linking assessment evidence to local action to improve children's learning.

These rich presentations informed participants' discussions in the e-Forum, and have been archived via YouTube for future reference. Please see the following link for access to the video playlist:

https://www.youtube.com/playlist?list=PLMoS8Y2K9JMduRfx6zbnpJMYsojRvb59f

E-Forum Conclusions

Participants clearly appreciated the opportunity to participate in this global e-Forum on the future of learning assessments at the national and international levels. While they initially discussed these two themes separately, over the course of the e-Forum several common issues of interest emerged. Overall, there were seven main points that summarize the results of this discussion:

1. *There is significant disagreement over whether an international assessment framework could be applicable to all countries*. Some participants argued strongly that a common global set of competencies could and should be measured, in order to be able to determine where more effort and resources might be needed to improve educational quality, and also because students' skills need to be internationally competitive. Others argued that countries have such different challenges and priorities, reflected in diverse curricula, that an international learning metric would be inadequate and misleading.

2. *There are differing opinions on which competencies should be assessed, in either a national or international context*. In general, most participants felt that assessments need to take into account more than just the basic academic skills of literacy and numeracy, although they agreed that these skills provide the foundation for learning. They argued that assessment should be expanded to address other competencies such as problem-solving, creativity, intrapersonal, and interpersonal skills, with others also adding vocational and entrepreneurship skills to the list. However, it was also acknowledged that these competencies can be very difficult to measure, and that cross-cultural comparisons may not be valid.

3. *There is uncertainty about the role that digital literacy and technology should play in assessments*. A number of participants remarked that digital literacy is becoming just as important as traditional literacy and numeracy—though not everyone agreed that it should be a subject of testing. The use of computer-based tests is also controversial.

4. *Participants were concerned with how international emphasis on learning assessments might affect national curricula and teacher practices*. There was concern that overemphasis on assessments could narrow the teaching and learning experience, and many felt that assessments cannot fully measure the breadth and diversity of what they considered a valuable education. Yet some participants argued that it is appropriate for an international assessment to have some influence on national curriculum since there are certain global problems, such as climate change, about which all young people should be educated.

5. *Developing effective assessments requires input from the regional, national, and even local levels*. On the one hand, many participants agreed that it should be possible to look at the needs of certain countries in order to modify international assessment tools accordingly, although certain assessment topics may be so culturally-linked that complete modification might be necessary. In contrast, several participants thought that new international assessments should be devised through a bottom-up process in which each nation participates in determining what should be tested and how. At the national level, some also thought that students, community members, teachers, and schools should be involved in determining the assessment approach.

6. *The most appropriate unit of analysis for an assessment depends on the assessment's purposes*. Participants understood that current international assessments use the nation as the unit of analysis to compare education systems, and in order to look at the effects of differences in national policies and practices. However, participants also argued that it is important to have

assessment data at the school level, because school leaders and teachers are those most immediately able to take remedial action to improve learning outcomes.

7. *There is a need to build capacity—not just at the national or regional level, but also among teachers and school leaders themselves—to understand assessment approaches and accurately interpret and use assessment data*. At the national level, e-Forum participants believed that implementing an international assessment could help countries build more sophisticated technical and managerial capacities for undertaking large-scale assessments. However, some questioned whether even very technically capable countries have been able to effectively use assessment data to improve learning outcomes. For this reason, participants argued that education decision-makers as well as teachers need training in how to interpret assessment results and use them to inform new educational strategies.

Several participants concluded the discussion by looking forward to the next e-Forum to be held in 2016. As Eileen Nkwanga (Zambia/UK) wrote, assessment alone is not enough to improve learning outcomes: "Thank you for the opportunity to participate in this e-Forum. May I stress the need to ensure that the conditions for effective teaching and learning at the local level are improved to ensure improvements in achievement? I would hope that a future e-Forum would address this."

With thanks to all of the participants for their comments and ideas, the topic for the next IIEP Learning Portal e-Forum will be announced in early 2016.