

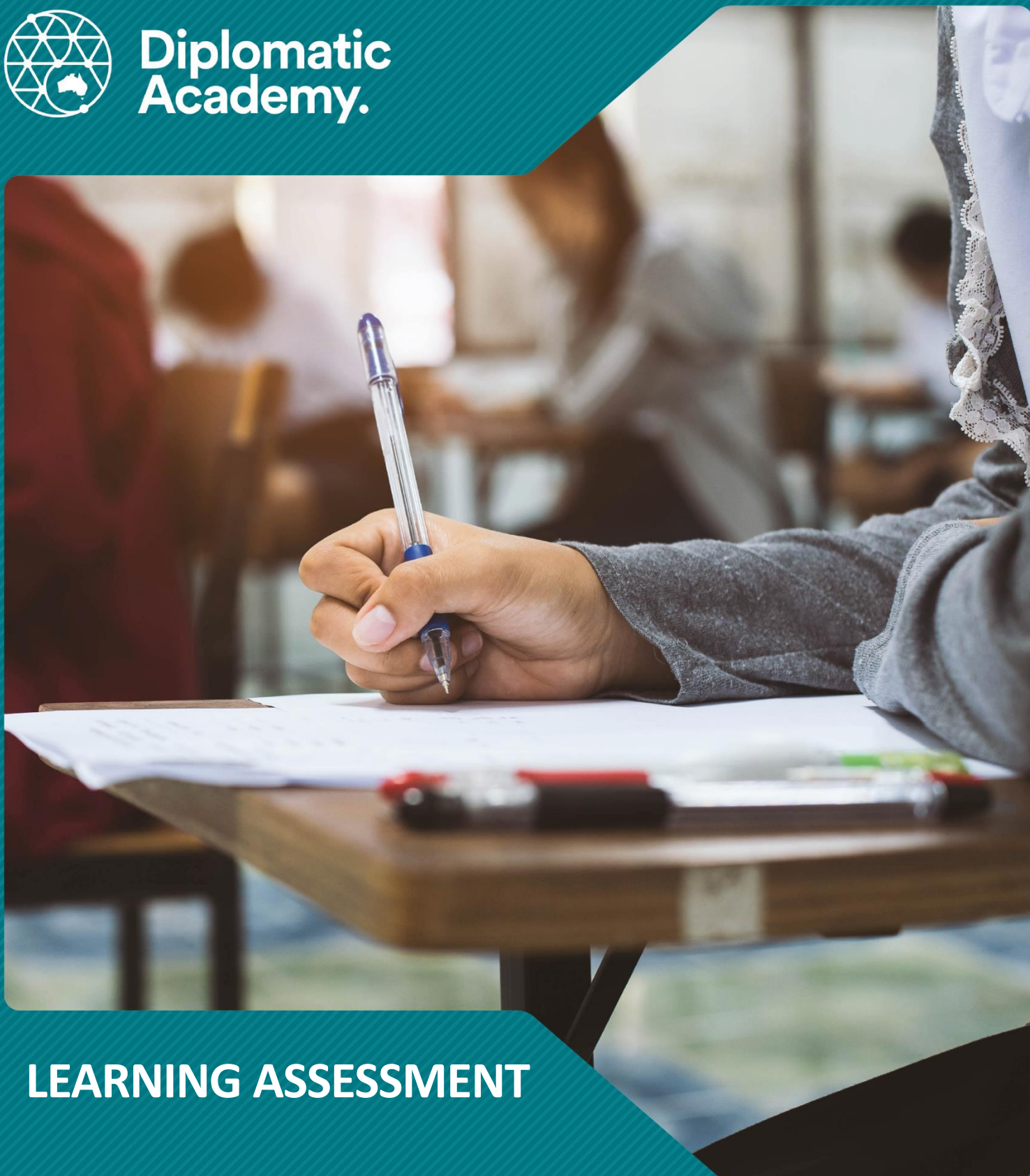


Australian Government
Department of Foreign Affairs and Trade

Education Learning and Development Module



**Diplomatic
Academy.**



LEARNING ASSESSMENT

Practitioner Level
2018



CONTENTS

Acronyms.....	3
1 Introduction	4
2 The development of learning assessment.....	4
3 Participating in international or regional assessments: benefits and riskS.....	6
4 National assessments: key design and implementation issues.....	7
5 Measuring and reporting results from assessment programs	9
6 The costs related to assessment programs	10
7 Assessment programs within developing countries.....	10
8 Inclusivity in learning assessment	12
9 Test your knowledge.....	13
References and links.....	15

Source: Cover photo, DFAT Education Section.



ACRONYMS

ACER	Australian Council for Educational Research
ASER	Annual Status of Education Report (India)
CTT	Classical Test Theory
DFAT	Australian Government Department of Foreign Affairs and Trade
DFID	Department for International Development (United Kingdom)
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
ICILS	The International Computer and Information Literacy Study
IEA	International Association for the Evaluation of Educational Achievement
IRT	Item Response Theory
LLECE	Latin American Laboratory for Assessment of the Quality of Education
MIA	Medición Independiente de Aprendizajes
NAPLAN	National Assessment Program – Literacy and Numeracy
NGO	non-government organisation
OECD	Organisation for Economic Co-operation and Development
PILNA	Pacific Islands Literacy and Numeracy Assessment
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
SACMEQ	The Southern and Eastern Africa Consortium for Monitoring Educational Quality
SEA-PLM	Southeast Asia Primary Learning Metrics
TIMSS	Trends in International Mathematics and Science Study
UIS	UNESCO Institute of Statistics
UNESCO	United Nations Educational, Scientific and Cultural Organization



1 INTRODUCTION

This Practitioner level module is designed to ensure that all staff members who engage with and lead policy dialogue with international and domestic partners are informed about the benefits and risks of participating in learning assessments and the key design and implementation issues.

It is recommended that staff complete the *Learning Assessment: Foundation level* module as background information to this Practitioner level module.

2 THE DEVELOPMENT OF LEARNING ASSESSMENT

What is a learning assessment?

Learning assessments are used to measure student achievement in order to identify strengths and areas for improvement.

The last 20 years have seen an increase in the number of learning assessments conducted, by both developed and developing countries, with greater emphasis on the importance of making evidence-based policy and practice decisions. This has resulted in many countries developing their own learning assessment programs, often with support from development partners.

Source: Benavot & Köseleci 2015, *Seeking quality in education: The growth of national learning assessments: 1990 –2013*.

International and national assessment program choice

Many countries conduct a range of assessment programs, including national and regional/international assessments. Different types of assessments can provide countries with different information that can be used to inform policy and practice decisions. In summary, national assessments are generally conducted within a country, and the data is typically used to measure student performance against expected performance (for example, against curriculum standards).

Many national assessments incorporate some questions from international assessments such as the Program for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS). A country must apply for permission to use released items from international assessment programs.

Many assessment programs also collect background or contextual information about factors that may be associated with achievement. This information is often collected through questionnaires. In international assessments, some questionnaire items are adapted so that they are suitable for the different countries participating.



International, national and regional assessments

There are countries that participate in international assessments and who also conduct their own national assessments. For example, Australia carries out its own National Assessment Program: Literacy and Numeracy (NAPLAN) and also participates in the following international assessment programs: PISA, TIMSS, the Progress in International Reading Literacy Study (PIRLS), and the International Computer and Information Literacy Study (ICILS).

Some countries participate in regional assessments such as the Southeast Asia Primary Learning Metrics (SEA-PLM) and the Pacific Island Literacy and Numeracy Assessment (PILNA). These include countries in a particular region that may share some socioeconomic, cultural and/or language features which can offer more possibilities for relevant cross-country comparisons.

The benefits and risks of different types of assessments are discussed later in this module.

An activity for you

SEA-PLM is a regional assessment program to help improve learning of all students in the Southeast Asia region.

Refer to this link to [SEA-PLM](#) and answer the following questions:

Question 1: What domains (subject areas) are included in SEA-PLM?

Question 2: What age students are assessed in SEA-PLM?

Check your answers.

Question 1: SEA-PLM includes these domains: reading, writing, mathematics and global citizenship.

Question 2: Grade 5 students are assessed in SEA-PLM.

Source: Southeast Asia Primary Learning Metrics (SEA-PLM) n.d, About SEA-PLM.





3 PARTICIPATING IN INTERNATIONAL OR REGIONAL ASSESSMENTS: BENEFITS AND RISKS

Key benefits and risks of international and regional assessments

There are many considerations when countries are deciding whether to participate in various international and/or regional assessments. Table 1 summarises some key benefits and risks.

Table 1: International and regional assessment benefits and risks

Potential Benefits	Potential Risks
Adherence to high technical standards of assessment test design, instrumentation, sampling, administration, analysis and reporting meaning that robust assessment data is collected.	Countries that are new to international/regional assessments may find it challenging to meet rigorous technical standards, including timelines. There is also a high financial cost of participation particularly considering that participation should continue over successive rounds if the initial investment is to be worthwhile and to gain useful trend data.
Development of local capacity to meet international standards of assessment practice. Opportunities for enhanced collaboration between different stakeholders involved in the assessment.	Disaffection with the international exercise if the assessment is of limited relevance to a country's needs.
Provides information to countries which can help with comprehensive monitoring of their education system. This includes providing information on trends over time and information on outcomes for different sub-groups. The results can be used by policy makers to inform decisions, including those related to policy reforms, resource allocation and programs.	The assessment results may not be well-utilised. Media reports often over-simplify assessment results solely focusing on country-rankings. Many countries experience political challenges when results are viewed as less positive than expected.
Publically accessible databases provide opportunities for secondary analyses of the data by researchers and other interested stakeholders, which can lead to further insights into the education system.	Assessments may be of limited value if the assessment items are not an appropriate level of difficulty for the population. International assessments are taking steps to expand the range of items included in order to address this.
Opportunity to adapt international survey instruments to the national language and context.	Only certain adaptations are permitted to the survey instruments. Significant time is required to ensure that the instruments are appropriate for the national context, and are adequately adapted and translated

Sources: Kellaghan, Bethell & Ross 2011, National and international assessments of student achievement, Table 3, p.17; Ho 2013, Student learning assessment.



4 NATIONAL ASSESSMENTS: KEY DESIGN AND IMPLEMENTATION ISSUES

What to assess?

In general, national assessments measure learning achievements in the areas of **literacy** (language) and **numeracy** (mathematics). Sometimes other areas such as scientific knowledge or social studies are included. National assessments typically focus on testing what students have learned and understood, and use a combination of multiple choice and ‘constructed response’ questions. Constructed response questions require the test taker to produce the answer (e.g. free text or short answer questions), rather than selecting from a list (e.g. multiple choice question).

The costs related to marking constructed response questions are much higher than for multiple choice items. This is because multiple choice questions can be machine marked, while constructed response questions need human raters to carry out marking. Finding the balance between the two (cost; accurately assessing learning) is a critical decision for national assessments. To accurately assess performance, a range of question types are typically used.

There are two main sampling approaches:

- **Population-based assessments:** an assessment involving all students and schools assessed at a specific age/grade. Australia’s National Assessment Program – Literacy and Numeracy (NAPLAN) is an example, with all students in Years 3, 5, 7 and 9 participating.
- **Sample-based assessments:** an assessment involving a sample that is representative of the age/grade. PISA is an example, with a representative sample of 15 year olds of participating countries sitting for the assessment.

At what stage to assess?

Each international and regional assessment varies in the population assessed. Some assess a particular age group, for example PISA assesses 15 year olds, while others assess particular grade levels. For example, PILNA assesses Year 4 and Year 6 students.

Source: Greaney & Kellaghan 2008, *Assessing national achievement levels in education*, p.31.

For national assessments, a country must decide at what stage students will be assessed. Clear policy goals are needed to guide decisions about who and what to assess. For example, a country may be interested in assessing student achievement and growth in student achievement between mid-primary school, upper primary school and early secondary school.

In developing countries, there is often much variation in the age at which students start school and it can be common for students to repeat one or more grades. In these circumstances there is a strong argument for testing by grade, rather than age.



Most assessments are designed to be administered in schools. However, where there are large-populations of out-of-school students there have been some organisations that have developed assessments designed to be administered in the community in order to capture these populations. For example, the [Annual Status of Education Report \(ASER\)](#) conducted in India is a household survey that includes the assessment of out of school students. It is run by a non-government organisation (NGO) and is sometimes referred to as a citizen-led assessment.

ASER is also run in Pakistan. Other citizen-led assessments include Beekunko in Mali, Jàngandoo in Senegal, Uwezo in Kenya, Tanzania, and Uganda, Medición Independiente de Aprendizajes (MIA) in Mexico, and LEARNigeria in Nigeria. See, for example, [Bringing learning to light: The role of citizen-led assessments in shifting the education agenda](#).

Source: Results for Development Institute 2015.

There may also be practical considerations when thinking about what students to assess. For example, students in the senior grades often have examinations and therefore, a decision might be made not to have a national monitoring assessment at the same time. If younger students are assessed, there will need to be particular considerations around how the tests will be administered. For example the Early Grade Reading Assessment (EGRA) and the Early Grade Mathematics Assessment (EGMA) are administered one-on-one for Grades 1 to 3 students.

The Early Grade Reading Assessment

EGRA is a diagnostic tool, typically used to assess student literacy in Grades 1 to 3. EGRA is used to gain an understanding of age-appropriate literacy levels, to ‘diagnose’ local conditions, and to thereby suggest appropriate actions. EGRA can be used to establish literacy benchmarks and targets and to assess literacy progress over time.

EGRA contains tasks and sub-tasks that measure the following processes of reading: phonological awareness, phonics, vocabulary, fluency and comprehension. The EGRA tool assesses children in their country language in a one-on-one oral situation. EGRA was not designed to measure cross-national data as the instruments can be adapted for different countries and languages.

EGMA is a diagnostic tool that measures students’ acquisition of basic mathematics skills. Like EGRA, it is an oral assessment that is administered one-on-one in schools. EGMA includes the assessment of number identification, reasoning about magnitude, recognition of number patterns, addition and subtraction and word problems.

The frequency of assessment

The frequency of assessments will vary by the assessment tool, by country and in relation to the intended purpose. An ideal frequency depends upon the assessment objectives.

For example, Greaney and Kellaghan state:

...if the aim of an assessment is to hold teachers, schools and even students accountable for their learning, testing may be carried out every year... If the purpose of an assessment is only to provide information on the performance



of the system as a whole, however, an assessment of a sample of students in a particular curriculum area every three to five years would seem adequate.

They also discuss the need to consider the cost implications of conducting frequent assessments.

Source: Greaney & Kellaghan 2008, *Assessing national achievement levels in education*, p. 43.

5 MEASURING AND REPORTING RESULTS FROM ASSESSMENT PROGRAMS

Classical Test Theory and Item Response Theory

Most achievement testing has, until recently, analysed data using **Classical Test Theory (CTT)**, with results presented as the proportion of correct answers. This scoring system is readily comprehensible to teachers and parents. Such an approach, however, is specific to particular tests and groups of students, and limits comparisons between assessments and over time.

Item Response Theory (IRT) has a number of advantages over CTT and has become the standard analytical technique for major national and international programs. IRT uses a measurement model to estimate the probability of a test item of a particular difficulty being answered correctly by a student of a given ability level. IRT also allows for the difficulty level of a test item to be described independently of a student who responds to the test item. Likewise, a student's achievement level is described independently of the group of items provided to the student. In large-scale assessments it is common for different groups of test items to be given to different students. If two students of the same ability are provided with two different tests – one easier test and one harder test – it is important that their achievement level is described independently of the test difficulty. IRT is therefore particularly helpful when different groups of items are given to students so that scores can be compared between tests. However, it is important that if IRT is used, there is sufficient expertise in using these complex procedures.

For further information see [Analysing Data from a National Assessment of Educational Achievement](#).

Source: Shiel & Cartwright 2014.

The analysis and reporting for assessments depends on the assessment design and purpose. For example, some assessments are designed to measure the learning outcomes of groups and sub-groups within a population (for example, the outcomes for boys and girls or of students from rural or urban areas). Assessments may also be designed to measure progress over time of an individual or groups of learners. This can be helpful for teachers to understand and address the changing needs of students.



6 THE COSTS RELATED TO ASSESSMENT PROGRAMS

The costs of assessment programs vary widely, and will depend on factors such as the type of assessment, the assessment frequency, the population assessed, and the resources and capacity available. Costs will also vary considerably between countries.

For any assessment it is essential that sufficient budgetary planning is conducted. Assessments are also most useful when they provide information over time. Therefore, it is important that sufficient resources are allocated to support assessments over a long time period. Greaney and Kellaghan (2012) provide a budget checklist in their handbook to give countries an overview of the different assessment stages that should be budgeted for in a national assessment.

As outlined by the authors, it is important that the budget also includes sufficient provision for capacity building, where needed, salary increases, inflation and contingency for unexpected events.

Source: Greaney & Kellaghan 2012, *Implementing a national assessment of educational achievement*

For more information about examples of costs for particular national, regional and international assessments, see the UNESCO-UIS paper [The cost of not assessing learning outcomes](#).

While this paper highlights that there are significant costs associated with implementing assessments, it also draws attention to the costs associated with not assessing students. High-quality assessments can provide important information about issues of quality and equity within an education system, with the ultimate aim of improving student learning outcomes. The authors also point to the relationship between educational performance and economic growth and state that “it is clear that it is very expensive to not assess education, with the consequent risk of not providing students with the skills demanded by the labour market and increasing inefficiencies”.

Source: UNESCO-UIS 2016, p. 15.

7 ASSESSMENT PROGRAMS WITHIN DEVELOPING COUNTRIES

A systematic review carried out by Best et al. (2013) investigated the impact of national and international assessment programmes on educational policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries. The report focused on 54 studies in developing countries. Most of the assessments included in the review used a sample-based approach, rather than a population or census-based approach.

The review identified that assessment data was most likely to be used in educational policies related to curriculum standards and reform, performance standards, and assessment



policies. However, there was little information about the impact of assessment programs on teaching and learning practices, or their impact on policy formation.

Source: Best et al 2013, The impact of national and international assessment programs on educational policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries.

Key barriers to successful assessment programs

The review outlined that some of the barriers to the utilisation of assessment program data included the quality of programs, financial restrictions and low technical capacity. In some cases the absence of adequate analysis of ‘associated factors’ such as age, gender, socioeconomic status and other background information means that the data is insufficiently anchored in the specific context.

Another review discussed some of the facilitators to the utilisation of assessment results. For example, assessment data is more likely to be used if it is integrated into the policy cycle, where the assessment program is of high quality, and where the information is appropriately disseminated to stakeholders. The authors also looked specifically at the impact of large-scale assessments on policy in the Asia-Pacific region, where they provided the following overall recommendations:

- Strive for integration of large-scale assessments in policymaking processes.
- Work to improve the technical quality of assessments, including developing the capacity of those involved in their design and implementation.
- Ensure that assessments have a sound communication and dissemination strategy that engages all relevant stakeholders in this effort, including the media.

Source: Tobin et al. 2015, Using large scale assessments of students’ learning to inform education Policy: Insights from the Asia-Pacific region, pp. 12-13.

As noted above, one issue identified was the lack of adequate analysis of factors such as socioeconomic status and the relationship with achievement. Subsequent to this review, there has been further work conducted to investigate measures of socioeconomic status that are appropriate for use in low and middle income countries. Information about forthcoming papers on this work are available at these links:

- 🔗 [Towards a global scale of household wealth](#)
- 🔗 [Quality of SES measures](#)
- 🔗 [Measures of SES in economics.](#)

Sources: Dix et al. 2017, p. 55; Adams et al. 2017a, p. 55, 2017b, p. 54.



8 INCLUSIVITY IN LEARNING ASSESSMENT

When designing assessment methods and tasks, it is essential that the methods used are appropriate for the domains (subjects) being assessed and are also appropriate for gathering information about all learners that are the focus of the assessment.

One important area to consider is that assessment tasks are inclusive. That is, the methods and tasks should be equitable for all students. Bias in assessments is where an assessment gives an advantage or disadvantage to a group or groups of students. For example, if the content of a test is of more interest to boys than girls, then this may result in gender bias. Or if a topic is more familiar to students in a particular region or students from a particular ethnic group, this may give advantages to these students.

In national, regional or international assessments, the question development process should include thorough review of assessment methods to ensure that they are inclusive of all students. The questions should also be trialed and statistically analysed, for example to see if there is any gender bias.

In the case of classroom assessment, it is important that teachers are conscious of the way in which assessment contexts and methods may advantage and disadvantage certain students. For example, topics that may cause students to feel distressed should be avoided.

In summary, “an assessment process should not provide underestimates or overestimates of some students’ levels of achievement or progress because of their gender, physical disability, cultural background or geographical location.”

Source: Masters G 2013, *Reforming education assessment: Imperatives, principles and challenges*, p. 41.



9 TEST YOUR KNOWLEDGE



Assessment questions

Answer the following questions by ticking 'True' or 'False'. Once you have selected your answers to all the questions, turn the page to 'The correct answers are...' to check the accuracy of your answers.

Question 1

National assessments are generally run within the country, but may incorporate data from international assessment programs.

Is this statement true or false? ☐ True ☐ False

Question 2

National assessments use both multiple-choice and constructed response items. Multiple choice questions are more expensive to mark.

Is this statement true or false? ☐ True ☐ False

Question 3

In developing countries there is a preference for testing students by grade, rather than by age.

Is this statement true or false? ☐ True ☐ False

Question 4

The Early Grade Reading Assessment (EGRA) is a diagnostic tool, not a high-stakes test.

Is this statement true or false? ☐ True ☐ False



The correct answers are...



Question 1

National assessments are generally run within the country, but may incorporate data from international assessment programs.

This statement is true. Many national assessments include questions from, for example, the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS).

Question 2

National assessments use both multiple-choice and constructed response items. Multiple choice questions are more expensive to mark.

The statement is false. Constructed response items are more expensive to mark because they need a human rater to interpret the written responses. Multiple-choice questions can be machine marked.

Question 3

In developing countries there is a preference for testing students by grade, rather than by age.

The statement is true. Grade level testing is often preferred because it can be easier to administer (e.g. class-based testing) and because of potentially large within-grade age ranges (e.g. due to the variation in starting age and with students who repeat a grade).

Question 4

The Early Grade Reading Assessment (EGRA) is a diagnostic tool, not a high-stakes test.

The statement is true. EGRA is used to gain an understanding of age-appropriate literacy levels, to 'diagnose' local conditions, and to thereby suggest appropriate actions.



REFERENCES AND LINKS

All links retrieved September, 2018

Adams, R, Ahmed, S K, Cloney, D, Dix, K, Friedman, T, Lietz, P, Routitsky A and Schwantner, U 2017a, 'Quality of SES measures' (session abstract), paper presented at the 7th IEA International Research Conference, 28-30 June 2017, Prague, Czech Republic, http://www.iea.nl/sites/default/files/publications/Electronic_versions/IRC%202017%20Program.pdf

2017b, 'Measures of SES in economics, education, health and psychology' (session abstract), paper presented at the 7th IEA International Research Conference, 28-30 June 2017, Prague, Czech Republic, http://www.iea.nl/sites/default/files/publications/Electronic_versions/IRC%202017%20Program.pdf

Benavot, A & Köseleci, N 2015, Seeking quality in education: The growth of national learning assessments: 1990 –2013, Background paper prepared for the EFA Global Monitoring Report 2015, Education for All 2000–2015: Achievements and challenges, UNESCO, <http://unesdoc.unesco.org/images/0023/002337/233733e.pdf>

Best, M, Knight, P, Lietz, P, Lockwood, C, Nugroho, D & Tobin, M 2013, The impact of national and international assessment programs on educational policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries, final report, Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre), Social Science Research Unit, Institute of Education, University of London, https://research.acer.edu.au/cgi/viewcontent.cgi?article=1016&context=ar_misc

Dix, K, Friedman, T, Cloney, D, Lietz, P, Routitsky, A, Schwantner, U, Ahmed, S K 2017, 'Towards a global scale of household wealth' (session abstract), paper presented at the 7th IEA International Research Conference, 28-30 June 2017, Prague, Czech Republic, http://www.iea.nl/sites/default/files/publications/Electronic_versions/IRC%202017%20Program.pdf

Greaney, V & Kellaghan, T 2008, Assessing national achievement levels in education, National Assessments of Educational Achievement, vol. 1, World Bank, <http://hdl.handle.net/10986/6904>

2012, Implementing a national assessment of educational achievement, National Assessments of Educational Achievement, <http://hdl.handle.net/10986/2237>

Ho, E S 2013, Student learning assessment, Asia-Pacific Education System Review Series, no. 5, <http://unesdoc.unesco.org/images/0021/002178/217816E.pdf>

Kellaghan T, Bethell, G and Ross, J 2011, National and international assessments of student achievement, Guidance Note: A DFID Practice Paper, <https://www.gov.uk/dfid-research-outputs/assignment-report-guidance-note-national-and-international-assessments-of-student-achievement>

Masters, G 2013, Reforming education assessment: Imperatives, principles and challenges, Australian Education Review, no. 57, <https://research.acer.edu.au/aer/12/>



Results for Development Institute 2015, Bringing learning to light: The role of citizen-led assessments in shifting the education agenda, <https://www.hewlett.org/wp-content/uploads/2016/08/R4D%20-%20Bringing%20Learning%20to%20Light%20-%20June%202015.pdf>

Shiel, G & Cartwright, F 2015, Analysing data from a national assessment of educational achievement, National Assessments of Educational Achievement, vol. 4, <https://openknowledge.worldbank.org/handle/10986/21433>

Southeast Asia Primary Learning Metrics (SEA-PLM) 2017, SEA-PLM About us, <http://www.seaplm.org/seaplm/>

Tobin, M, Lietz, P, Nugroho, D, Vivekanandan, R, & Nyamkhuu, T 2015, Using large-scale assessments of students' learning to inform education policy: Insights from the Asia-Pacific region, September, ACER and UNESCO Bangkok, <http://unesdoc.unesco.org/images/0023/002354/235469e.pdf>

UNESCO-UIS 2016, The cost of not assessing learning outcomes, Information Paper No. 26, January, <http://dx.doi.org/10.15220/978-92-9189-184-9-en>

Learn more about...

- ✂ PISA, found at: <http://www.oecd.org/pisa/>
- ✂ TIMSS and PIRLS, found at: <https://timssandpirls.bc.edu/>
- ✂ NAPLAN, found at: <http://www.nap.edu.au/naplan/naplan.html>
- ✂ The Southern and East African Consortium for Monitoring Educational Quality (SACMEQ), found at: <http://www.unesco.org/education/pdf/SACMEQ.PDF> and <http://research.acer.edu.au/cgi/viewcontent.cgi?article=1007&context=assessgems>
- ✂ The Latin American Laboratory for Assessment of the Quality of Education (LLECE), found at: <http://www.unesco.org/new/en/santiago/education/education-assessment-llece/>
- ✂ The Early Grade Reading Assessment (EGRA), found at: <https://globalreadingnetwork.net/resources/early-grade-reading-assessment-egra-toolkit-second-edition>
- ✂ The USAID analysis of EGRA and gender, found at: https://ierc-publicfiles.s3.amazonaws.com/public/resources/EGRA_gender_brief_11Oct2011_FINAL-rev1_EndDateUpdated-Jun2014.pdf
- ✂ The 2011 Early Grade Reading Assessment (EGRA) applications and interventions to improve basic literacy, found at: <http://www.rti.org/pubs/bk-0007-1109-wetterberg.pdf>
Assessing National Achievement Levels in Education at this website: <http://hdl.handle.net/10986/6904>
- ✂ The impact of national and international assessment programs on educational policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries in a report by Knight, Lietz, Nugroho & Tobin (2012), found at: <http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=7E5NLbtPMPc=&tabid=3174>



- ✎ *The national, regional and international learning assessments trends in developing countries*, found at: <http://dx.doi.org/10.1080/14767724.2011.577337>
- ✎ *The Malawi National Early Grade Reading Assessment*, found at: https://globalreadingnetwork.net/sites/default/files/eddata/EGRA_2011_midterm_report.pdf and http://pdf.usaid.gov/pdf_docs/PA00JB9Q.pdf
- ✎ *A new conceptual framework for analyzing the costs of performance assessment* by Picus, Adamson, Montague & Owens (2010), found at: <https://scale.stanford.edu/system/files/new-conceptual-framework-analyzing-costs-performance-assessment.pdf>
- ✎ *How effective student assessment systems can help achieve learning outcomes* by Clarke (2012), found at: <https://openknowledge.worldbank.org/bitstream/handle/10986/10058/673460revised000110AssessmentOPress.pdf?sequence=1>
- ✎ *Comparing the similarities and differences of PISA and TIMSS*, found at: <http://dx.doi.org/10.1787/5km4psnm13nx-en>
- ✎ *The early grade reading assessment: Applications and interventions to improve basic literacy* edited by Gove and Wetterberg (2011), found at: https://www.researchgate.net/publication/280156924_The_Early_Grade_Reading_Assessment_Applications_and_Interventions_to_Improve_Basic_Literacy
- ✎ *National, regional and international learning assessments: Trends among developing countries, 1960–2009*, by Kamens & Benavot (2011), found at: <https://www.tandfonline.com/doi/abs/10.1080/14767724.2011.577337>
- ✎ *National assessments of educational achievement* by Postlethwaite & Kellaghan (2008), found at: <http://unesdoc.unesco.org/images/0018/001817/181753e.pdf>
- ✎ *Comparing the similarities and differences of PISA 2003 and TIMSS* by Wu, found at: <http://dx.doi.org/10.1787/5km4psnm13nx-en>

Wolff, L 2007, *The costs of student assessment in Latin America*, working paper 38, Partnership for Educational Revitalization in the Americas (PREAL), Washington, DC.