

CENTRE FOR
EDUCATION
POLICY AND
PRACTICE

CENTRE FOR
GLOBAL
EDUCATION
MONITORING

Using large-scale assessments of students' learning to inform education policy

Insights from the Asia-Pacific region

M. Tobin, P. Lietz, D. Nugroho,
R. Vivekanandan and T. Nyamkhuu

SEPTEMBER 2015



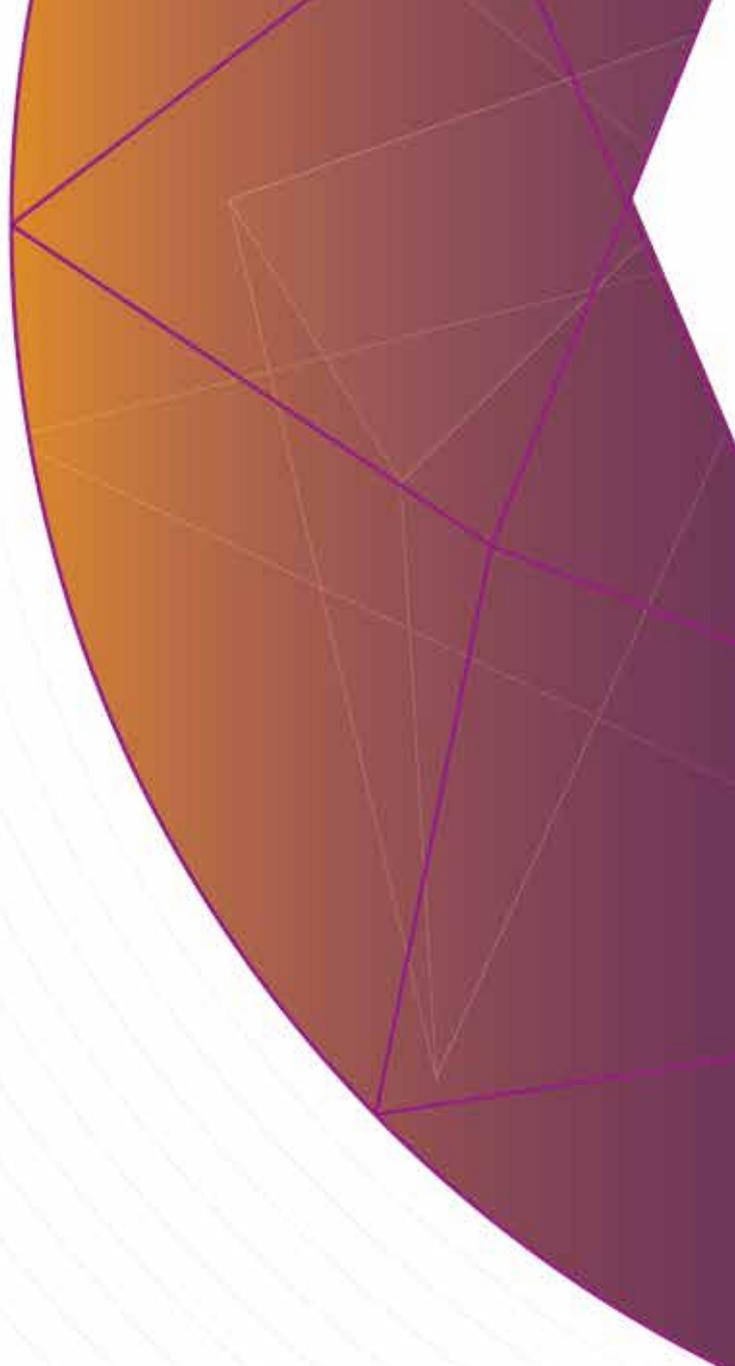
United Nations
Educational, Scientific and
Cultural Organization

Bangkok Office
Asia and Pacific Regional Bureau
for Education



NEQMAP
Network on Education Quality
Monitoring in the Asia-Pacific





Tobin, M.¹, Lietz, P.¹, Nugroho, D.¹, Vivekanandan, R.², & Nyamkhuu, T.² (September 2015). Using large-scale assessments of students' learning to inform education policy: Insights from the Asia-Pacific region. Melbourne: ACER and Bangkok: UNESCO.

1 Australian Council for Educational Research
2 UNESCO Bangkok

Published by
Australian Council for Educational Research
19 Prospect Hill Road
Camberwell VIC 3124
Phone: (03) 9277 5555

www.acer.edu.au

Australian Council for Educational Research © 2015

ISBN: 978-1-74286-296-5

All rights reserved. Except under the conditions described in the Copyright Act 1968 of Australia and subsequent amendments, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the written permission of the publishers.

www.acer.edu.au/epp
www.acer.edu.au/gem

Design: ACER Creative Services
Printed by: On-Demand



CONTENTS

Overview	2
Introduction	2
What types of assessments have influenced education policy in the region?	3
What are the intended uses of assessments?	4
How do policymakers use assessment data?	6
What education policies have been informed by assessments?	8
What factors influence the use of assessments in education policy?	10
Where to from here?	12
Conclusion	14
References	15
Abbreviations	17
Countries in the Asia-Pacific region for which evidence was found in the review	17

OVERVIEW

This paper presents results from a systematic review of literature that examined the link between participation in large-scale assessments of students' learning and education policy in the Asia-Pacific region.

The review was conducted by the Australian Council for Educational Research (ACER) through its Centre for Global Education Monitoring (GEM). It was a joint activity with the Network on Education Quality Monitoring in the Asia-Pacific (NEQMAP), for which the United Nations Educational, Scientific and Cultural Organization (UNESCO) Bangkok serves as Secretariat. NEQMAP is a regional platform on student learning assessment that supports the capacity development of those implementing and/or coordinating large-scale assessments in the Asia-Pacific region, particularly in low- and middle-income countries.

INTRODUCTION

In the late 1950s, cross-national large-scale assessments were conceived with the aim of using countries as natural laboratories to explore student learning (Foshay, 1962). In addition, many countries have a long history of using examination systems to certify individual student learning or to select students for further study.

After the establishment of Education for All (EFA) in 1990, there has been rapid growth in the number of countries participating in large-scale assessments of students' learning, particularly in low- and middle-income countries. EFA is a global movement led by UNESCO to coordinate development efforts across countries, institutions and other organisations to work towards meeting education goals for all children and youth (UNESCO, 2015).

Large-scale assessments of students' learning:¹

- ▶ are standardised to enable comparability across students, schools and in some cases, countries
- ▶ are intended to be representative of an education system either at the sub-national (i.e., state, province) or national levels
- ▶ are equally likely to be conducted in centralised or decentralised education systems

- ▶ in some instances can compare education systems across countries in the same region² or internationally³
- ▶ *do not* have as their main purpose to certify individual student achievement, and do not refer to assessments used by teachers in classrooms, or to selective or 'gate-keeping' assessments such as graduation examinations or university entrance examinations.

Countries of all income levels in the Asia-Pacific region are increasingly likely to have participated in a large-scale assessment of students' learning. Benavot and Köseleci (2015) highlight that by 2013, 69 per cent of countries in the region had carried out a national assessment. This compares with only 17 per cent in the 1990s. Examining the global growth of national assessments, close to a quarter of all national assessments undertaken around the world between 2007 and 2013 were conducted in the Asia-Pacific region.

This growth in participation has been accompanied by a shift in the use of assessments, from the exploration of differences between education systems to the evaluation of education service delivery and outcomes (Kamens & McNeely, 2009). Assessments are intended to provide information for evidence-based policy and decision-making about education inputs and resourcing, with a view to the continuous improvement of learning outcomes.

Concerns continue to be raised about the usefulness of international assessments for policymaking (Goldstein & Thomas, 2008) and the use of national high-stakes assessments. Nevertheless, policy- and decision-makers are reinforcing the use of assessments to monitor progress towards education development goals for the 2030 education agenda (UNESCO, 2015) and documenting country participation in assessment activities (UNESCO Institute for Statistics, 2015).

Still, not much is known about the ways in which assessment data have actually been used in education policy to date. Understanding the role of assessments in informing system-level decision-making is a first step towards helping stakeholders improve the design and usefulness of

1 The term 'assessment' is used in this paper to refer to large-scale assessments of students' learning.

2 For example: Pacific Islands Literacy and Numeracy Assessment (PILNA); the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ); Conference of the Ministers of Education of French Speaking Countries' (CONFEMEN) Programme for the Analysis of Education Systems (PASEC); the Latin American Laboratory for Assessment of the Quality of Education (LLECE).

3 For example: Programme for International Student Assessment (PISA); Trends in International Mathematics and Science Study (TIMSS); Progress in International Reading Literacy Study (PIRLS).

assessments. Moreover, this understanding can help to further discussions about how assessment data can best be used to inform policy and practice and to evaluate the effectiveness of policy reforms.

This paper presents results from a systematic review of 68 studies that examined the link between participation in large-scale assessment programs of students' learning and education policy in 32 countries in the Asia-Pacific region.⁴ Included studies either identified specific cases of assessment results being used by policymakers to inform education reform in their systems, or identified specific cases when assessment results had no impact on education policy in specific education systems. The review classified the available evidence to address the questions:

- ▶ What types of assessments have impacted education policy in the region?
- ▶ What are the intended uses of assessments?
- ▶ How are assessment data used in education policy?
- ▶ What education policies have been informed by assessments?
- ▶ What factors influence the use of assessments in education policy?

WHAT TYPES OF ASSESSMENTS HAVE INFLUENCED EDUCATION POLICY IN THE REGION?

Evidence of large-scale assessments of students' learning being used in education policy was primarily found in literature about Australia, Japan, New Zealand, India, Indonesia and Singapore. Even though many low- and middle-income countries in the Asia-Pacific region are undertaking national assessments or participating in regional or international assessments, much less is known about the role assessments play in education policy in these contexts. Figure 1 illustrates the frequency of evidence included in the review by country.

Assessments that have been found to impact on education policy are more frequently:

- ▶ national rather than international assessments
- ▶ assessments at secondary rather than primary school level
- ▶ sample-based rather than population (census)-based assessments.

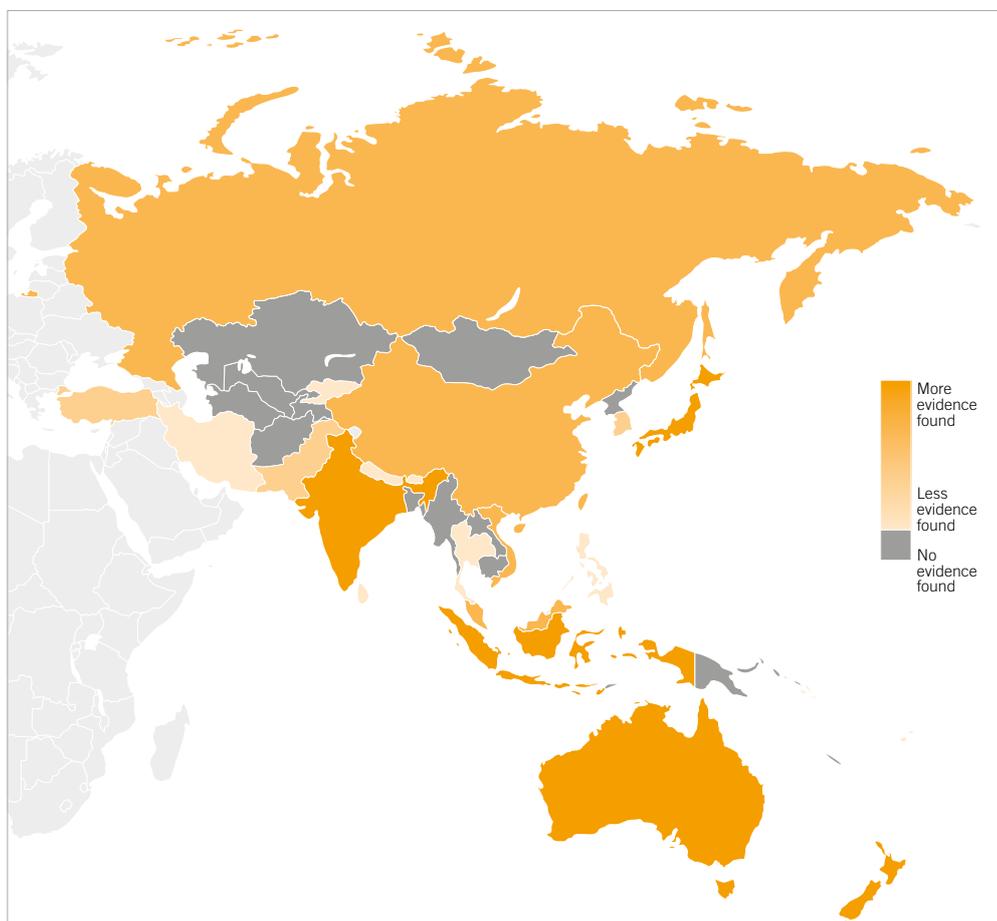


Figure 1: Evidence of impact of assessments on education policy in the Asia-Pacific region by country.

⁴ Asia-Pacific countries for which the review found evidence are listed at the end of the paper.

WHAT ARE THE INTENDED USES OF ASSESSMENTS?

Quality



While large-scale assessments of students' learning are often used for multiple purposes, the assessment programs that are linked to policy in the Asia-Pacific region are more frequently intended to ensure the *quality* of the education system. These assessments diagnose system strengths and weaknesses over time through system monitoring.

JAPAN USED the Programme for International Student Assessment (PISA) and the Japanese national assessment program to develop an 'evidence-based improvement cycle' to monitor the quality of its education system over time (Wiseman, 2013). The Japanese Ministry of Education (MEXT) was able to identify a suite of issues for education reform through monitoring Japan's performance in PISA over time, from 2000 to 2009. This monitoring was complemented by the concurrent identification of issues through Japan's national assessment program, starting in 2007. In order to improve the targeting and implementation of the identified issues for reform, MEXT developed an improvement cycle to specify how reforms would be implemented and monitored at the national, local and school levels (Suzuki, 2011).

After quality, assessments are equally intended to ensure *equity* of the education system for subgroups, and *accountability* of the education system for improving students' learning outcomes.

Equity



Assessments can be used to ensure equity of the education system by examining education outcomes for specified subgroups. Subgroups of interest are often those which have historically experienced educational disadvantage, such as girls, children in rural and remote areas, or children from low-socioeconomic backgrounds. Assessments can monitor outcomes for these subgroups, and inform initiatives that aim to address educational inequity.

AUSTRALIA'S PARTICIPATION in international assessments, such as PISA, has been used to monitor achievement differences between students from different socioeconomic backgrounds (Dinham, 2013). The country's national assessment program has been used to monitor achievement differences between Indigenous and non-Indigenous students (Ford, 2013).

Accountability



Assessments can also be used for accountability purposes, with the aim of improving educational quality and equity by reporting assessment outcomes to stakeholders internal or external to the education system.

National assessments, and the few sub-national assessments included in this review, are more often associated with accountability goals than are international assessments. In addition, assessments that use a census to test all students in an education system at specified year levels are more frequently associated with accountability goals than are sample-based assessments.

SOUTH KOREA reintroduced its national assessment program in 2008, to be used as an accountability tool. The national assessment program had been discontinued from 1998 to 2007, but in 2008 the new government instituted an annual National Diagnostic Exam, a census assessment of all students in year 3, and a National Curriculum Exam of all students in years 6, 9 and 10. Aggregate results are reported to internal stakeholders such as schools and the federal government. Results are also reported to external stakeholders, primarily the media and parents, so that teachers and schools can be held accountable for students' learning (Sung & Kang, 2012).

Leverage

Some of the literature in this review that is critical of the relationship between assessments and policy argues that assessment programs are sometimes used to *leverage* pre-existing political priorities. The goal of leverage is least frequently mentioned in the literature, in comparison to the goals of quality, equity and accountability. Yet, when this review considered literature that did mention the use of assessments to leverage political priorities, it found that participation in international assessments is more frequently mentioned in association with leverage than other assessment types.

For example, assessments can provide 'external policy support' with the public and other stakeholders (Gür, Zafer, & Özo lu, 2012) to prioritise a government's particular education reform agenda. Using assessments to leverage political priorities is in contrast with using assessments to consider an education system's context in an evidence-based policy approach. Figure 2 below summarises findings about the intended goals and uses of assessments in the Asia-Pacific region.

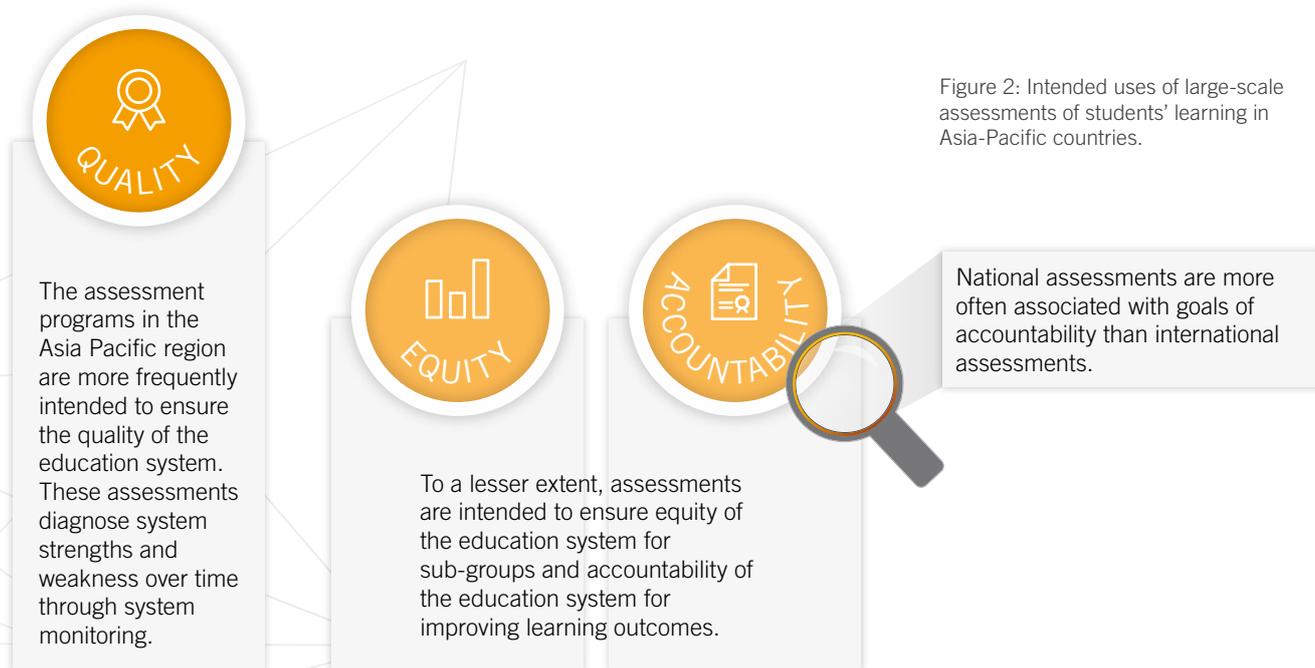


Figure 2: Intended uses of large-scale assessments of students' learning in Asia-Pacific countries.

HOW DO POLICYMAKERS USE ASSESSMENT DATA?

Education policy may be understood as policy change at one or numerous stages of a policy cycle. This review used a simplified model of a policy cycle (Sutcliffe & Court, 2005), which separates policymaking into four stages: agenda-setting, policy formulation, policy implementation, and monitoring and policy evaluation. Large-scale assessments of students' learning can be considered by policymakers at one or more stages of the policy cycle.

Monitoring and evaluation

Assessments are most frequently used by policymakers to *monitor and evaluate* education policies, and in the development of monitoring mechanisms. National assessments are used more frequently for monitoring and evaluation purposes, in comparison to international assessments. The monitoring and evaluation stage of the policy cycle considers the establishment of monitoring mechanisms to provide information, and processes to evaluate implemented policies or initiatives. This stage of the policy cycle intends to provide information about a policy outcome to inform future or ongoing decision-making.

VIETNAM HAS used national assessment results to monitor students' learning outcomes over time, in order to help evaluate the effectiveness of policy initiatives for improving educational quality. Vietnam has conducted a national assessment of year 5 students in reading and mathematics in 2001, 2007 and 2011. In parallel with these assessment cycles, Vietnam implemented the Primary Education for Disadvantaged Children (PEDC) project (2004–2010), which targeted resource allocation and service delivery in disadvantaged areas to help schools meet new school-based standards, or the Fundamental School Quality Level (FSQL). Vietnam has used results to evaluate specific policies of the PEDC and FSQL initiatives, including the implementation of a new curriculum, teacher-student contact hours, and implementation of new school-based standards (Attfield & Vu, 2013).

The development of monitoring mechanisms frequently refers to the use of assessment results to create or reform monitoring and evaluation units and to legislate evaluation activities. For example, in 2008, the first year of the Australian national assessment program, results were used by some state education departments to establish units for the further monitoring and examination of students' learning outcomes at the state-level (Lingard & Sellar, 2013).

Policy implementation

The second most frequent use of assessment results by policymakers is for *policy implementation*. This stage involves the use of evidence from assessments to improve the effectiveness of the ways in which an initiative is targeted or implemented on the ground.

Most often, policy implementation refers to the use of assessments in implementing curricular or programmatic reforms.

IN THE Indian state of Madhya Pradesh, the education department established a state-wide Learn to Read initiative in 2005, in order to improve student literacy outcomes. Standardised student assessment results were used from 2006 to 2010 to support the implementation of the Learn to Read initiative. The data allowed the provision of teacher coaches and other supports to be effectively targeted to districts, schools and teachers. The education department also used standardised student assessment data to target additional remuneration for teachers (Mourshed, Chijoke, & Barber, 2010).

Agenda-setting

Assessments are also equally likely to be used for *agenda-setting*. Policymakers in the Asia-Pacific region use assessment results to create awareness and give priority to an issue for reform, most often the quality of some aspect of the education system. Assessment results are also used at the agenda-setting policy stage to create awareness about the magnitude of an identified issue.

International assessments are used by policymakers in the agenda-setting stage to evaluate the quality of the education system through the comparison of students' learning relative to other countries.

RUSSIA'S PARTICIPATION in several international assessments from 1995 to 2011 helped raise concerns about perceived declining educational quality, particularly at the secondary level of education. In addition, Russia's continued participation in international assessments helped to raise policymakers' awareness about the importance of the social and school contexts for students' learning. Raising decision-makers' awareness over time ultimately led to reform of the curriculum and performance standards at both primary and secondary levels of education (Bolotov, Kovaleva, Pinskaya, & Valdman, 2013; Tyumeneva, 2013).

Policy formulation

Assessments are used least frequently for *policy formulation*, which refers to the design and formulation of policy options and the selection of a policy strategy. High-income countries in the Asia-Pacific region are more likely than low- or middle-income countries to use assessments during this stage of the policy cycle.

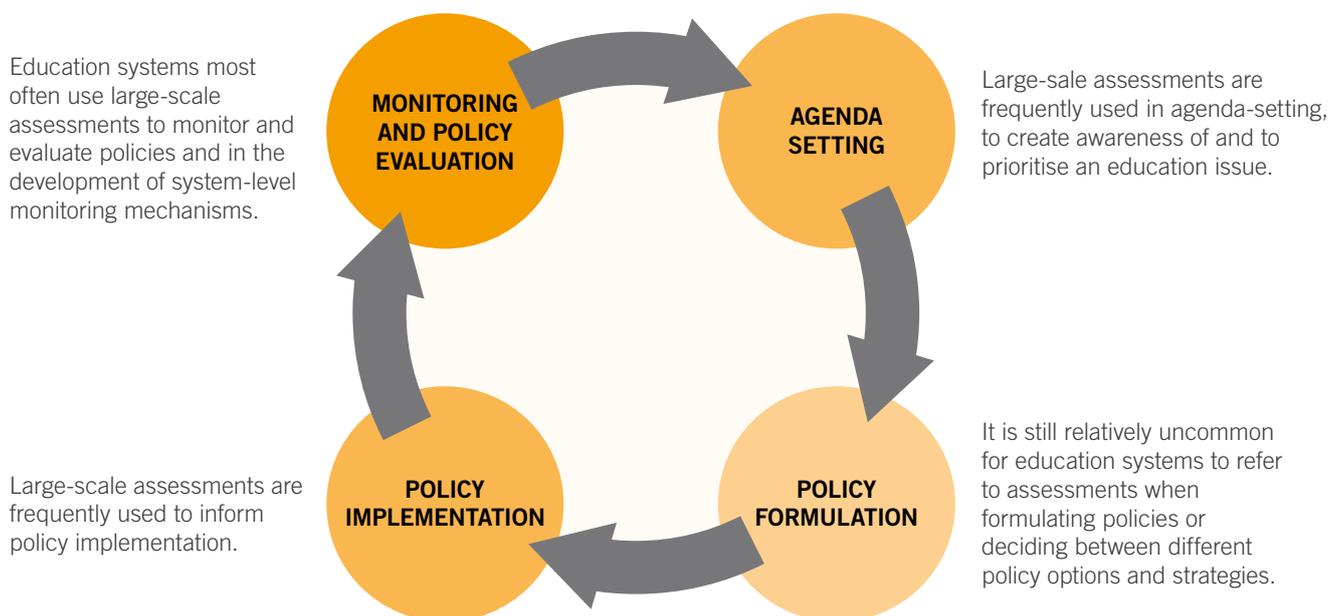
THE RELEASE of PISA results in 2001 showed that Japan had not performed as well as expected in reading literacy. Consequently, policymakers in Japan legislated an Act in 2002, the Fundamental Plan for Promotion of Reading, which required all students in lower and upper primary levels to participate in a daily morning reading session. Further decline in PISA results in 2003 prompted policymakers to legislate more comprehensive policies to target the improvement of student literacy in both 2005 and 2006 (Ninomiya & Urabe, 2011).

In some instances, studies note that assessments have had little to no impact on education policy in Asia-Pacific countries. This means that while education systems conduct or participate in an assessment, results are not used by policymakers for education decision-making. This was reported across all assessment types, including sub-national, national, regional and international assessments. A closer examination of instances of no impact shows that barriers to the use of assessment data in education policy include:

- ▶ perceived low technical quality of the assessment program
- ▶ lack of in-depth and policy-relevant analyses to be able to identify and diagnose issues
- ▶ poor timing of the assessment program and non-integration of the assessment into policy processes
- ▶ inappropriately tailored dissemination to stakeholders
- ▶ lack of dissemination to the public.

Figure 3 summarises how large-scale assessments are used in the education policy cycle.

Figure 3: Use of large-scale assessments in the education policy cycle.



WHAT EDUCATION POLICIES HAVE BEEN INFORMED BY ASSESSMENTS?

This review classified education policies according to a framework that broadly grouped specific policies as system-level policies, resource allocation policies, or teaching and learning policies. Figure 4 illustrates the review's results whereby large-scale assessments most frequently influence system-level policies, followed by resource allocation policies. Large-scale assessments least frequently affect policies which directly impact on teaching and learning in classrooms.

System-level policies



Large-scale assessments of students' learning are most frequently used to inform *system-level policies*, which provide a framework for evaluation systems and operations. These

include assessment policies and policies regulating curricular and performance standards.

Assessments are most frequently used to inform the development of *assessment policies* for the further monitoring and evaluation of the education system. These policies often establish or modify the conduct and use of assessments at system and local levels.

IRAN'S PARTICIPATION in the Trends in International Mathematics and Science Study (TIMSS) led to the use of the TIMSS curricular framework in the development of test items for the country's own assessment uses (Heyneman & Lee, 2014).

The establishment and reform of curricular and performance standards aim to provide a common framework and context for the interpretation of assessment results.

KYRGYZSTAN'S LOWER than anticipated results in PISA 2006 helped to support and inform the government's ongoing curricular reforms in 2009 and 2010, to emphasise 'modern skills and competencies' in line with those assessed by PISA, and to foster an expectation of higher student performance (Shamatov & Sainazarov, 2010).

PISA RESULTS helped to inform the development of student performance standards in Japan's 'New Growth Strategy' in 2010, to be achieved by 2020. The performance standards outline goals for academic achievement. The standards also include expectations for higher proportions of students to report positive attitudes and interest towards learning, which was highlighted in recent PISA results (Breakspear, 2012).

Resource allocation policies



After system-level policies, assessments most frequently influence *resource allocation policies*, which refer to the ways in which resources are determined and allocated within an education

system. In the Asia-Pacific region, assessments are most often used to influence resource allocation policies targeting *in-service professional development programs* and *instructional materials*.

In this review, in-service professional development policies can refer to changes in the focus, delivery or frequency of professional development programs to stakeholders such as school leaders and teachers.

AUSTRALIA USED national assessment data to target in-service professional development programs to identified schools through a National Partnerships initiative to promote the improvement of teacher and school quality. In-service professional development programs included providing literacy and numeracy coaches to work with targeted school staff for improvement of pedagogy (Council of Australian Governments, 2012).

Resource allocation policies may target pedagogical or instructional materials such as textbooks or other teaching resources.

NEW ZEALAND'S Ministry of Education developed a series of books to improve teachers' knowledge and teaching of basic science concepts in primary education, after perceived poor results in TIMSS (Jones & Bunting, 2013).

Teaching and learning policies



There is less available evidence of large-scale assessments having an impact on *teaching and learning policies*, which are aimed at specific school- and classroom-level practices. Teaching policies, as conceptualised in the review, may relate to factors such as: classroom management, differentiated teaching and support for students, professional collaboration and learning, teacher-student relationships, job satisfaction and efficacy. Learning policies, as conceptualised in the review, may consider factors such as: enhanced learning activities, collaborative or competitive learning, and programs to support students' interest and motivation in school.

In instances where such a link between assessments and teaching and learning policies was evident, the impact frequently was on policies for *in-class learning strategies*.

SHANGHAI'S CURRICULAR reform, which commenced in 1998 and is ongoing, intends for teachers to implement more student-centred pedagogies and promote students' learning through 'participation, real-life experience, communication and teamwork, and problem-solving'. Shanghai's 2009 PISA results provided the Shanghai Municipal Education Commission with evidence that the new teaching and learning policies associated with the ongoing curricular reform have been successful and should continue (Tan, 2012).

Teaching and learning policies also target *enhanced learning strategies* that require higher order thinking skills.

MALAYSIA FOCUSED on improving learning activities in science lessons by increasing the frequency of using experiments and computers after participating in TIMSS in 2003 (Gilmore, 2005).

Some of the literature argues that these tests have intended or unintended impacts on teaching and learning in classrooms, rather than on any legislated policy at a system level. Most often, these critiques highlight a narrowing of teacher-implemented curriculum to align more closely with what is assessed in these large-scale assessment programs (Klenowski & Wyatt-Smith, 2012).

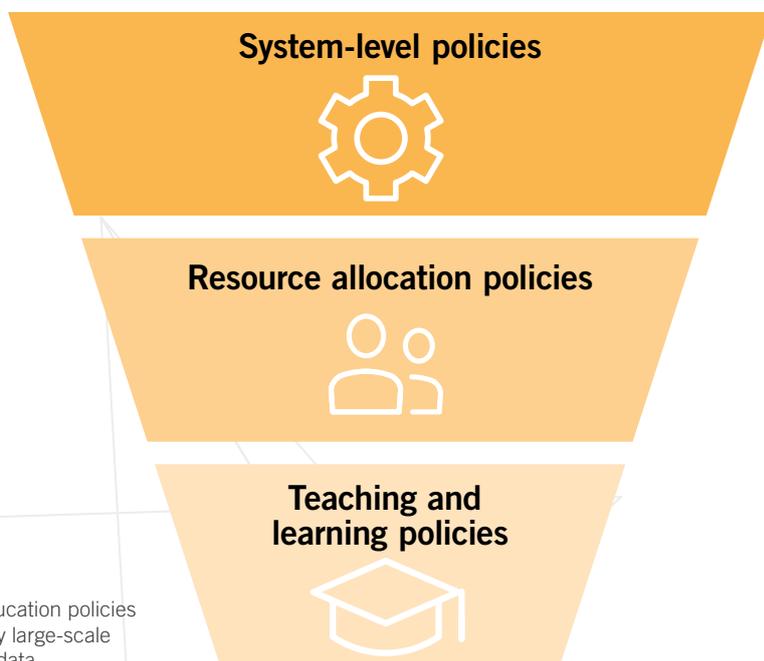


Figure 4: Education policies influenced by large-scale assessment data.

WHAT FACTORS INFLUENCE THE USE OF ASSESSMENTS IN EDUCATION POLICY?

This review identified a number of factors that influence the use of large-scale assessments in education policy in Asia-Pacific countries. These factors were found to facilitate or inhibit the use of assessments. Overall more facilitators than barriers were found to influence the use of assessments in education policy in the Asia-Pacific region.

Integration into policy processes



Integration into policy processes is the most frequently cited factor influencing the use of assessment data in education policy. Legislating assessment programs provides a legal mandate for the regular conduct and financing of assessments and a platform for their use in education policy. Assessment agencies that are long-term and well-funded help the assessment agency to remain insulated from political instability and regime change, and for results to be considered seriously by stakeholders and the public. Assessment agencies that are mandated by government have greater authority in the policy process and when responding to government's policy priorities than assessments that have no mandate, or are 'one-off'.

Assessments that are external to the education system, such as large-scale citizen-led assessments, are also able to integrate into policy processes by aligning the assessment goals and reporting in part with policy priorities and legislation.

PAKISTAN'S CITIZEN-LED household assessment, the Annual Status of Education Report (ASER) Pakistan, has aligned its assessment goals and reporting with monitoring progress towards government-legislated development priorities, thereby increasing its use for government reporting and monitoring in education (ASER, 2014).

Media and public opinion



The influence of the *media and public opinion* is a key factor affecting the use of assessment results. Often, international and regional assessments receive a great deal of media attention and the publication of high-level results can create a 'policy window' through which to place the issue of educational quality on the policy agenda.

RESULTS FROM Malaysia's participation in TIMSS Repeat (TIMSS-R) in 1999 made front-page news, with media coverage of the subsequent parliamentary debate about educational quality. This coverage helped in part to inform the government's renewed emphasis on science and mathematics education. The media coverage and dissemination of TIMSS-R results also helped to influence the government's decision to participate in subsequent cycles of TIMSS (Elley, 2005).

To a lesser extent, some literature characterises media coverage as a barrier to the development of meaningful or effective policies, by increasing pressure on policymakers to adopt politically attractive policies and quick solutions (Lingard & Sellar, 2013). On the other hand, a lack of dissemination and media coverage, due to political sensitivities, can act as a barrier to the use of assessment results in policymaking (Attfield & Vu, 2013; Levine, 2013).

Quality of the assessment program



The *quality of the assessment program* can be either a facilitator or a barrier to the use of assessment results in education policy. While the technical quality of programs is more frequently cited as a facilitator for international assessments, it is more frequently considered a barrier for national assessments. In some instances, participation in an international assessment supported a country's technical capacity development to undertake a national assessment. For example, the methodologies applied in international assessments were then used by stakeholders in Russia in the development of Russia's national assessment program (Bolotov et al., 2013).

Poor technical quality of an assessment may limit the use of data in education policy. For example, technical concerns related to sampling and field operations may affect the perceived representativeness and legitimacy of a survey's results in the eyes of stakeholders (Kellaghan, Bethell & Ross, 2011).

The inability to diagnose policy-relevant issues and undertake in-depth analyses also serves as a barrier to its use in education policy. For example, technical issues related to the non-comparability of assessment cycles over time (Maligalig & Albert, 2008) may make it difficult for policymakers to use assessment results to monitor trends or evaluate policy initiatives or interventions.

This review identified that integration into policy processes, media and public opinion, and the quality of the assessment program influence the use of large-scale assessments of students' learning.

WHERE TO FROM HERE?

The findings of this review are a step towards developing an understanding not only of the ways in which large-scale assessments of students' learning are being used to inform education policy, but also of the factors that influence their use. This paper highlights specific factors that countries in the Asia-Pacific region can consider to improve the design and use of assessments in evidence-based education policy.

Strive for integration of large-scale assessments in policymaking processes

Integration into policy processes was cited as an important factor that has influenced the use of assessment results in education policy. Integration includes legislating assessment programs in order to provide a legal mandate for the regular conduct and financing of assessments and a platform for their use in education policy. By adopting such legislation, assessment agencies themselves are more likely to be long-term in orientation and adequately funded, thereby gaining a measure of insulation from political instability and regime change and also increasing the legitimacy of the assessment agency and results with external stakeholders and the public. Still, efforts to integrate assessments in policy processes have to be cognisant of the perception of the independence of assessment programs and implementation agencies.

In addition, assessments have to be recognised as part of the policy cycle by policymakers and practitioners. Stakeholders involved in education reform should seek to identify effective and appropriate ways for assessments to align with policy processes.

Assessments are most frequently used to inform system-level policies for the further monitoring and evaluation of the education system.

To this end, both policymakers and practitioners should be involved in key stages of assessments, including in the identification of policy-relevant issues in the initial design of the assessment, and in the analysis of assessment data, thereby facilitating the effective use of assessment results for education policy. With this aim, the following are suggestions that can be considered in order to improve the integration of assessments in education policy processes:

- ▶ Formally legislate the establishment, conduct and financing of assessment programs and agencies.
- ▶ Ensure that information relevant to identified policy concerns is obtained in the assessment.
- ▶ Include questions about factors related to student outcomes (e.g., students' socioeconomic background and availability of resources at school and at home).
- ▶ Organise regular meetings and seminars between officials responsible for conducting assessments and policymakers in order to facilitate communication and understanding of results.
- ▶ Ensure that the reporting of assessment results includes policy papers specifically targeted to policymakers, in accessible language and linking back to policy issues of concern.

Work to improve the technical quality of assessments, including developing the capacity of those involved in their design and implementation

The technical soundness of the assessment is an important factor that has influenced the relationship between assessments and policymaking. To design and maintain the quality of assessments, highly developed technical skills are required at all stages of the assessment, from design and development, sampling, test administration and data collection, data cleaning and analysis, and reporting and dissemination of results. Capacity building of stakeholders who are engaged in assessments is essential.

In addition, further work could support policymakers in understanding how issues related to the technical quality and analysis priorities have implications for the initial design and funding of the assessment. Ensuring that the technical quality of the assessment supports its intended purpose will help to

strengthen the usefulness of the data for decision-making. In this regard, the following measures can be considered to improve technical quality of assessments:

- ▶ Ensure that best practice is followed in the design and implementation of the assessment (Clarke, 2012).
- ▶ Consider engagement in international or regional assessment programs that emphasise national capacity building so that the technical skills of assessment staff may be applied to national assessment programs.
- ▶ Pursue capacity development opportunities for assessment agency staff and policymakers, including through regional networks, technical assistance agencies, university courses or other training programs.

Ensure that assessments have a sound communication and dissemination strategy that engages all relevant stakeholders in this effort, including the media

The media and dissemination of assessment results to the public were identified as important factors influencing the use of assessment results in education policy. Therefore, it is important to effectively disseminate and communicate assessment results not only to those directly involved in assessment programs but also to all relevant stakeholders. How results are reported and presented, and the timing of the release, has to be established. Montoya (2015) highlights that dissemination of assessment results is also strongly related to the purpose and use of the assessment. Therefore results should not just be disseminated in a general manner, but instead be targeted to different stakeholders engaged in education reform. Since various stakeholders, including parents, teachers and policymakers, are involved in education reform, it is worth giving consideration to the interests and technical knowledge of each stakeholder group and producing different reports based on the particular needs and interests of each, while supporting discussions about realistic timelines and options for reforming practice and policy.

The review noted that the influence of the media and public opinion can be an important facilitator to leverage the use of assessment results in education policy. More work could focus on identifying effective ways of engaging with and disseminating results to the media. This could enable the

media to be informed and effective partners in disseminating assessment results and communicating with the public in this regard. To this end, policymakers can consider the following suggestions:

- ▶ Ensure the dissemination of assessment results to all stakeholders and target dissemination according to the interests and technical knowledge of each stakeholder group (e.g., via different types of reports and forums for communication).
- ▶ Engage with the media through all phases of an assessment program in order to increase the media's understanding and facilitate better communication.

Large-scale assessments are less frequently used to inform teaching and learning policies, which aim to affect school- and classroom-level processes.

CONCLUSION

Overall, the available evidence that publicly examines the link between large-scale assessments of students' learning and education policy is limited. The reason for this might be that evidence for such links, for example in ministerial briefings, is likely to be confidential and not available for public scrutiny. The bulk of evidence that was found for this review comes from high-income countries in the Asia-Pacific region, from Australia, Japan and New Zealand.

Much less evidence was found regarding the ways that assessments feature in education policy in low- and middle-income countries in the region, even though these countries are increasingly likely to have participated in international assessments or conducted their own assessments. As low- and middle-income countries constitute the majority of countries in the Asia-Pacific region, the relationship between assessments and education policy should be further explored in these contexts to support evidence-based decision-making in the region, while acknowledging that political sensitivities around educational quality and governance often limit the public availability of such analyses and discussions.

In addition, few studies in this review examined factors external to the assessment or education system. External factors can have significant impact on the use of assessments for policy reform. External issues may be related to political or economic instability, for example. Stakeholders who are increasingly focusing on supporting education reform in conflict-affected

and fragile states should consider the ways that external factors impact the use of assessment to inform educational reform and evidence-based education policy.

This review has shown that assessments are most frequently used to inform system-level policies, which include assessment policies for the further monitoring and evaluation of the education system.

Assessments are less frequently used to inform teaching and learning policies, which aim to affect school- and classroom-level processes.

Similarly, Montoya (2015) notes that stakeholders primarily use assessment data 'to assess and manage education systems' rather than using assessment data as 'a rich source of information to directly address the needs of students'.

A more nuanced understanding of the realities of the policy process at international, national and local levels can help policymakers, educators and other stakeholders to more effectively leverage assessment results at appropriate stages of the policy cycle. In this way, assessment results can better support stakeholders in identifying effective levers that will support bottom-up or 'micro' reform in schools (Masters, 2014), in order to improve students' learning outcomes.

REFERENCES

- Annual Status of Education Report (ASER). (2014). *Annual status of education report: ASER Pakistan 2013 national*. Lahore, Pakistan: South Asian Forum for Education Development (SAFED).
- Attfield, I., & Vu, B.T. (2013). A rising tide of primary school standards: The role of data systems in improving equitable access for all to quality education in Vietnam. *International Journal of Educational Development*, 33(1), 74–87.
- Benavot, A., & Köseleci, N. (2015). Seeking quality in education: The growth of national learning assessments, 1990–2013. Paper commissioned for the EFA Global Monitoring Report 2015, *Education for All 2000–2015: Achievements and challenges*. Paris: UNESCO.
- Bolotov, V., Kovaleva, G., Pinskaya, M., & Valdman, I. (2013). *Developing the enabling context for student assessment in Russia*. Washington, DC: The International Bank for Reconstruction and Development, The World Bank.
- Breakspear, S. (2012). *The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance*. Paris: OECD.
- Clarke, M. (2012). *What matters most for student assessment systems: A framework paper*. Washington, DC: The World Bank.
- Council of Australian Governments (COAG). (2012). *National partnership agreement on literacy and numeracy: Performance report for 2011*. Sydney: COAG Reform Council.
- Dinham, S. (2013). The quality teaching movement in Australia encounters difficult terrain: A personal perspective. *Australian Journal of Education*, 57(2), 91–106.
- Elley, W.B. (2005). How TIMSS-R contributed to education in eighteen developing countries. *Prospects*, 35(2), 199–212.
- Ford, M. (2013). Achievement gaps in Australia: What NAPLAN reveals about education inequality in Australia. *Race, Ethnicity and Education*, 16(1), 80–102.
- Foshay, A.W. (1962). *Educational achievements of thirteen-year olds in twelve countries: Results of an international research project, 1959–61* (Vol. 4). Hamburg, Germany: UNESCO Institute for Education.
- Gilmore, A. (2005). The impact of PIRLS (2001) and TIMSS (2003) in low- and middle-income countries. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).
- Goldstein, H., & Thomas, S.M. (2008). Reflections on the international comparative surveys debate. *Assessment in Education: Principles, Policy & Practice*, 15(3), 215–222.
- Gür, B.S., Zafer, C., & Özo lu, M. (2012). Policy options for Turkey: A critique of the interpretation and utilization of PISA results in Turkey. *Journal of Education Policy*, 27(1), 1–21.
- Heyneman, S.P., & Lee, B. (2014). The impact of international studies of academic achievement on policy and research. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues and methods of data analysis* (pp. 37–72). Boca Raton, Florida: Taylor and Francis Group.
- Jones, A., & Bunting, C. (2013). International, national and classroom assessment: Potent factors in shaping what counts in school science. In D. Corrigan, R. Gunstone, & A. Jones (Eds.), *Valuing assessment in science education: Pedagogy, curriculum, policy* (pp. 33 – 53). Netherlands: Springer.
- Kamens, D.H., & McNeely, C.L. (2009). Globalization and the growth of international educational testing and national assessment. *Comparative Education Review*, 54(1), 5–25.
- Kellaghan, T., Bethell, G., & Ross, J. (2011). *National and international assessments of student achievement*. Guidance Note: a DFID Practice Paper. London: Department for International Development (DFID).
- Klenowski, V., & Wyatt-Smith, C. (2012). The impact of high stakes testing: The Australian story. *Assessment in Education: Principles, Policy and Practice*, 19(1), 65–79.
- Levine, V. (2013). *Education in Pacific island states: Reflections on the failure of 'grand remedies'* (Pacific islands policy no. 8). Honolulu, Hawai'i: East-West Center.
- Lingard, B., & Sellar, S. (2013). 'Catalyst data': Perverse systemic effects of audit and accountability in Australian schooling. *Journal of Education Policy*, 28(5), 634–656.
- Maligalig, D.S., & Albert, J.G. (2008). *Measures for assessing basic education in the Philippines* (Discussion paper series no. 16). Makati City, Philippines: Philippine Institute for Development Studies.
- Masters, G.N. (2014). Is school reform working? *Policy Insights, Issue 1*. Australia: ACER.
- Montoya, S. (2015, June 17). Why media reports about learning assessment data make me cringe. *World Education Blog, Education for All Global Monitoring Report*. <https://efareport.wordpress.com/2015/06/17/why-media-reports-about-learning-assessment-data-make-me-cringe/>
- Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better*. London: McKinsey and Company.
- Ninomiya, A., & Urabe, M. (2011). Impact of PISA on education policy: The case of Japan. *Pacific-Asian Education*, 23(1), 23–30.
- Shamatov, D., & Sainazarov, K. (2010). The impact of standardized testing on education in Kyrgyzstan: The case of the Programme for International Student Assessment (PISA) 2006. *International Perspectives on Education and Society*, 13, 145–179.
- Sung, Y.K., & Kang, M.O. (2012). The cultural politics of national testing and test result release policy in South Korea: A critical discourse analysis. *Asia Pacific Journal of Education*, 32(1), 53–73.
- Sutcliffe S., & Court, J. (2005). *Evidence-based policymaking: What is it? How does it work? What relevance for developing countries?* London: Overseas Development Institute.
- Suzuki, K. (2011). PISA survey and education reform in Japan: Construction of an evidence-based improvement cycle. Presentation at the 14th OECD Japan seminar, 'Strong Performers, Successful Reformers in Education'. <http://www.oecd.org/>

pisa/14thoecdjapanseminarstrongperformerssuccessful
reformersineducation.htm

Tan, C. (2012). The culture of education policymaking: Curriculum reform in Shanghai. *Critical Studies in Education*, 53(2), 153–167.

Tyumeneva, Y. (2013). *Disseminating and using student assessment information in Russia*. Washington, DC: The International Bank for Reconstruction and Development, The World Bank.

United Nations Educational, Scientific and Cultural Organization Institute for Statistics (UIS). (2015). Learning outcomes. <http://www.uis.unesco.org/Education/Pages/observatory-of-learning-outcomes.aspx>

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2015). *Incheon Declaration — Education 2030: Towards inclusive and equitable quality education and lifelong learning for all*. Paris: UNESCO.

Wiseman, A.W. (2013). Policy responses to PISA in comparative perspective. In H.D. Meyer & A. Benavot (Eds.), *PISA, power, and policy* (pp. 303–322). Oxford: Symposium Books.

ABBREVIATIONS

ACER	Australian Council for Educational Research
ASER	Annual Status of Education Report
COAG	Council of Australian Governments
CONFEMEN	Conference of the Ministers of Education of French Speaking Countries
EFA	Education for All
FSQL	Fundamental School Quality Level
GEM	Global Education Monitoring
LLECE	Latin American Laboratory for Assessment of the Quality of Education
MEXT	Japanese Ministry of Education
NEQMAP	Network on Education Quality Monitoring in the Asia-Pacific
PASEC	Programme for the Analysis of Education Systems
PEDC	Primary Education for Disadvantaged Children
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
TIMSS	Trends in International Mathematics and Science Study
TIMSS-R	Trends in International Mathematics and Science Study — Repeat
UIS	UNESCO Institute of Statistics
UNESCO	United Nations Educational, Scientific and Cultural Organization

COUNTRIES IN THE ASIA-PACIFIC REGION FOR WHICH EVIDENCE WAS FOUND IN THE REVIEW

Australia	Marshall Islands	Singapore
Bhutan	Micronesia, Federated States of	Solomon Islands
China, including Hong Kong	Nauru	Sri Lanka
Fiji	Nepal	Thailand
India	New Zealand	Tokelau
Indonesia	Pakistan	Tonga
Iran	Palau	Turkey
Japan	Philippines	Tuvalu
Kiribati	Republic of Korea	Vanuatu
Kyrgyzstan	Russian Federation	Vietnam
Malaysia	Samoa	



www.acer.edu.au