

C H A P T E R

9

## Assessment in context

DOI: <https://doi.org/10.56383/NFAL3758>

*This chapter should be cited as:*

*Verhoeven, L., Li, X., Ogunniran, M.O., Wu, R., Chakraborty, A., Vickers, E.A., Gupta, L. and Macintyre, T. (2022). 'Assessment in context' in Duraiappah, A.K., van Atteveldt, N.M., Borst, G., Bugden, S., Ergas, O., Gilead, T., Gupta, L., Mercier J., Pugh, K., Singh, N.C. and Vickers, E.A. (eds.) Reimagining Education: The International Science and Evidence Based Assessment. New Delhi: UNESCO MGIEP.*



**T**his chapter analyses educational assessment in relation to its social, cultural, economic and political context. We suggest that the worldwide focus on assessment of learning, partly driven by the influence of the OECD, is related to the idea that allowing ‘merit’, defined in terms of measurable learning outcomes, to determine life chances is an adequate guarantor of social justice. From this dominant perspective, education is also seen primarily as serving to form productive workers rather than enhance flourishing or promote truly sustainable and equitable societies. The latter goal instead calls for a holistic and humanistic approach to education and a less high-stakes model of assessment for and as learning. We suggest that some important aspects of education for child flourishing cannot easily, or appropriately, be subjected to quantitative assessment. It has been claimed that technological advances enable better assessment of higher-order skills with the potential to improve teaching, but we argue that any benefits to teachers and students remain unclear. Understanding the complex relationship between assessment and context requires analysis of public discourse on science and evidence, the influences that shape it, and the role of vested interests. We therefore emphasize the need to recognize how ideological beliefs, economic interests and political agendas drive assessment reforms around the world, stressing that inclusive and holistic approaches to assessment need to take proper account of the sociocultural and economic concerns of the communities they serve.

## Coordinating Lead Authors

Ludo Verhoeven

Xiaomin Li

## Lead Authors

Moses Oladele Ogunniran

Rongxiu Wu

Anya Chakraborty

Edward A. Vickers

Latika Gupta

Thomas Macintyre



# 9.1

## Introduction

In recent decades, there has been a discernible expansion in the range, scope and intensity of educational assessment, from individual assessment to national classroom monitoring and cross-national comparisons. Much of this expansion has been underpinned or legitimated by claims for the role of science and evidence in the creation, implementation, evaluation and validation of assessment instruments. Advancements in the fields of neuroscience and information technology have been portrayed as challenging established assessment practices, offering the promise

of an ‘assessment revolution’. Meanwhile, international organizations, such as the World Bank, UNESCO and the OECD, have become key players in promoting and delivering cross-national comparisons based on quantitative assessment of educational ‘outcomes’, and in advancing the application of new technology in this area.

Assessment is an inescapable and necessary feature of any education system, and indeed of the activity of learning itself. As we discuss below, it can take many different forms and perform

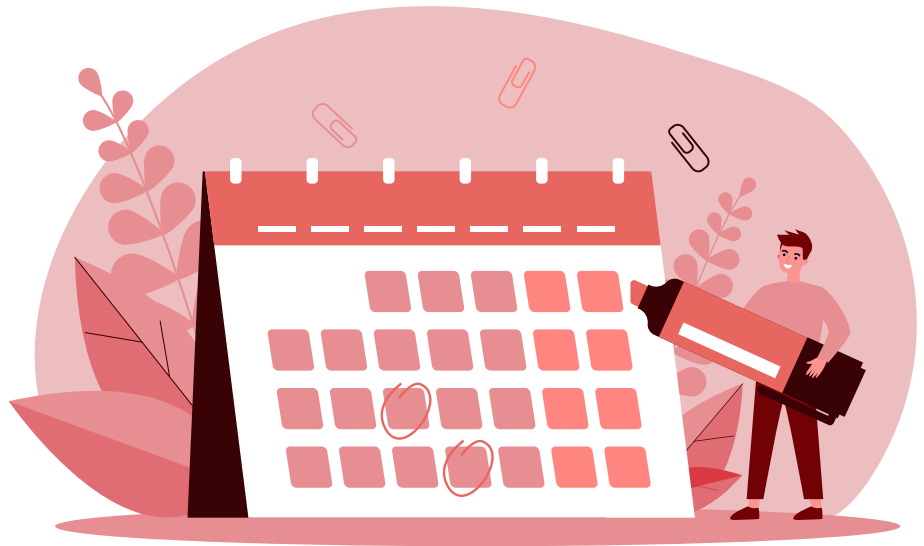


Assessment is an inescapable and necessary feature of any education system, and indeed of the activity of learning itself.

many different purposes. But assessment practices or techniques are far from the value-neutral, context-independent tools for the maximization of ‘effectiveness’ or ‘efficiency’ in learning and teaching that much public policy debate assumes them to be. As the previous chapter on curriculum and pedagogy emphasized, ideas concerning what should be taught and learnt, how, and why have always varied widely across space and time. This variation cannot be understood without grasping the ways in which the social, cultural and political context shapes educational institutions and beliefs. As with curriculum, so with assessment – our assumptions about what is worth assessing, why and how assessment should be conducted are fundamentally context-dependent. Assessment practices can serve to empower teachers and learners, but they can also disempower, narrowing the scope of learning and impoverishing education in the name of ‘accountability’.

This chapter seeks to explain this Janus-faced quality of educational

assessment in relation to the social, cultural and political context. Moreover, we also need to understand both the contextual influences on our assessment practices, and how those practices themselves form a crucial part of the context that shapes education. What does the drive to generate quantifiable metrics of educational ‘outcomes’ imply for the curricular status of fields, such as the arts and humanities, in which learning is especially difficult to measure quantifiably? Who should control assessment regimes or have access to the information they generate, and for what purposes? What is the relationship between educational assessment, ideology and social and political control? When and why does assessment of education become oppressive and intrusive, atomizing individual learners and reducing learning to little more than an intense competition to certify acquisition of epistemological and social ‘capital’? How much assessment is enough – or too much? And how do calls for an approach to assessment driven by ‘science and evidence’ relate to such questions?



# 9.2

## Assessment and learning

In this section, we review the forms of assessment, noting that educational assessment has served varied functions in different systems of education and at different times. Assessment has been entangled in structures of power embedded in political and economic regimes, ideas about capabilities entrenched in sociocultural hierarchies and ideological assumptions concerning ‘official knowledge’ (Apple, 1993). Therefore, assessment is not a neutral device; its aims and techniques are determined in complex ways by the context in which it operates.

**9.2 .1**

### FORMS OF ASSESSMENT

**9.2 .1 .1**

#### ASSESSMENT OF LEARNING (OR ‘SUMMATIVE ASSESSMENT’)

The summative assessment of learning can be traced back to the civil service examinations conducted by China’s imperial rulers during the first millennium CE. Its most basic function is to

rank and score learners, thereby lending meritocratic legitimacy to selection of individuals for opportunities for further study or employment. This form of assessment has been used to describe the process of evaluating the effectiveness of sequences of instructional activities to provide information for judging the overall value of an education programme – as well as for ranking and scoring learners for selective purposes. The process relates to the mechanics or steps required to effectuate a judgement, which cannot be made within a vacuum and require predetermined standards and goals to make comparisons.

In the specific area of assessing students' performance it is used to evaluate what students have learned at the end of a course or a grade level, or as a selection method for entry into educational tracks (William, 2011). A typical example of summative assessment today is the Gaokao system in China (modern descendent of the imperial civil service examinations of old) in which

the results of the annual national exam determine students' entrance to the higher education system or to access certain credentials such as an undergraduate degree or a technical job certificate (Gu, Ma and Teng, 2017). Data resulting from such assessment programmes can also report the quality of teaching, differences in achievement levels by subgroups (such as gender or region) and, if background data are collected, factors that contribute to reaching different levels of achievement (Lietz et al., 2008). School boards or ministries of education also often use summative assessments to keep publicly funded schools accountable for the provision of their education (Schildkamp and Kuiper, 2010).

9.2	.1	.2
-----	----	----

## ASSESSMENT FOR LEARNING (OR 'FORMATIVE ASSESSMENT')

In the latter half of the twentieth century, a considerable shift occurred from assessment of learning to assessment for

... has been used to describe the process of evaluating the effectiveness of sequences of instructional activities to provide information for judging the overall value of an education programme.

...‘assessment for learning’ stresses the ways in which assessment can contribute to efforts by teachers and students to seek, reflect upon and respond to information from dialogue, demonstration, and observation in ways that enhance ongoing learning.

learning to better capture the characteristics of assessment and to serve the purpose of furthering students’ learning (**Schuwirth and Vleuten, 2011**). Talk of ‘assessment for learning’ (or what is often dubbed ‘assessment’) stresses the ways in which assessment can contribute to efforts by teachers and students to seek, reflect upon and respond to information from dialogue, demonstration, and observation in ways that enhance ongoing learning (**Broadfoot et al., 2002**). Therefore, this form differs from the assessments designed primarily to serve the purposes of selection, accountability or certifying competence. It has an undoubted appeal when contrasted with the assessment of learning. However, the distinctions between assessment for learning and assessment of learning relate primarily to the purpose for which the assessment is carried out (**Black et al., 2004**). In practice, the same assessment techniques can often serve either a summative or formative purpose, depending on who is conducting the assessment, who has access to the resulting information, and how they choose to use it.

9.2

.1

.3

## ASSESSMENT AS LEARNING

Assessment as learning refers to the process whereby students are able to learn about themselves as learners and become aware of how they learn. Through self and peer assessment, students reflect on their progress and (often with the help of teachers) decide what their next learning objectives will be. It also helps students to take more responsibility for their own learning and monitoring future directions. The role of teachers in assessment as learning is to promote the development of independent learners by modelling and teaching the skills of self-assessment, guiding students in setting their own goals, and monitoring their progress towards them. Teachers also help create an enabling environment where students are encouraged to confront challenges and to develop strategies to adjust and advance their learning. It may be suggested that this form of assessment as learning associates with the idea of what Dweck (2008) calls a growth

mindset compared to a fixed mindset. Students with a growth mindset believe that their abilities can be cultivated and that success is about stretching themselves to learn something new.

## 9.2 .2

### OBJECTS OF ASSESSMENT

... the logic of human capital theory suggests that these domains are critical for preparing individuals to fulfil their productive potential, and thus contribute to maximizing (national or corporate) economic performance.

Whilst assessment as learning has great potential to empower students and foster ongoing learning, in contemporary education systems, the focus of assessment has been to identify and evaluate skills that are primarily related to the academic domains of reading, mathematics and science.<sup>1</sup> This has been driven fundamentally by the logic of human capital

theory (HCT), which suggests that these domains are critical for preparing individuals to fulfil their productive potential, and thus contribute to maximizing (national or corporate) economic performance (Hanushek and Woessmann, 2008). A classic example of this is the OECD's Program for International Student Assessment (PISA), a triennial assessment of students' performance in these key domains. Reluctant to be seen as advocating a crudely instrumentalist vision of education, statements of PISA's aims often conceal the economic focus that informs its overwhelming emphasis on maths, science and literacy. For example, mathematical literacy is described as comprising 'an individual's capacity to identify and understand the role mathematics plays in the world, to make well-founded judgements and to use and engage with mathematics in

<sup>1</sup>This contrasts starkly with the focus of assessment in ancient China, where imperial civil service examinations assessed mastery of a canon of classical texts dealing with philosophy, history and what today we might call 'public administration'. But in late imperial China, as around the world today, debate raged over appropriate ways to assess such learning, with many intellectuals harshly critical of the effects of a stilted, formulaic approach to assessment (the 'eight-legged essay') on cultural and political life.



ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen' (OECD, 2003, p. 24).<sup>2</sup> We discuss the OECD's assessment activities and their implications further below.

To validate assessment outcomes across populations of subjects, an attempt is often made to anchor the assessment items on a latent scale. As a case in point, scaling with the Rasch model or Item Response Theory is often used in cross-national or cross-cultural assessments to provide evidence of the fact that test items and student abilities can be anchored on the same underlying scale (Fischer and Molenaar, 2012; De Ayala, 2013). However, this idea has also been criticized as being part of a reductionist view of education (Wu et al., 2020). Cognitive diagnosis assessment (CDA) is another newly developed assessment theory that integrates the goals of assessment in the process of cognitive models and reflects students' potential cognitive

processes through their responses to items. It can effectively obtain students' advantages and deficiencies in fine-grained knowledge and provide a foundation for imparting students in accordance with their aptitude and adaptive learning (Wu et al., 2020).

In recent years, the OECD has been expanding the scope of PISA metrics to incorporate non-cognitive skills, such as social and emotional skills, global competency, creativity and well-being (e.g. Sellar and Lingard, 2014; Grotlüschen, 2018; Auld and Morris, 2019; Rappleye et al., 2019; Li and Auld, 2020). Since 2018, PISA has assessed 'global competence' (OECD, 2016a, p. 4), defined as 'the capacity to analyze global and intercultural issues critically and from multiple perspectives, to understand how differences affect perceptions, judgments, and ideas of self and others, and to engage in open, appropriate and effective interactions with others from

It can effectively obtain students' advantages and deficiencies in fine-grained knowledge and provide a foundation for imparting students in accordance with their aptitude and adaptive learning.

<sup>2</sup><https://www.oecd.org/education/school/programme-for-international-student-assessment-pisa/33707192.pdf>

... OECD has been attempting to position itself as the global arbiter for monitoring education systems' performance in relation to the United Nations Sustainable Development Goals, especially SDG 4.7.

different backgrounds on the basis of a shared respect for human dignity' (ibid). The declared aim is to evaluate the extent to which students are prepared to act in a globalized world in terms of their skills, attitudes and knowledge (Sälzer and Roczen, 2018). Here the OECD has been attempting to position itself as the global arbiter for monitoring education systems' performance in relation to the United Nations (UN) Sustainable Development Goals (SDGs), especially SDG 4.7: 'education for peace, sustainable development and global citizenship' (UNESCO MGIEP, 2017; Auld and Morris 2020). This effectively brings the OECD into competition with UNESCO, which has traditionally adopted a more humanistic rather than human capital-oriented perspective on education. But in the process of devising ways of assessing the SDGs, 'global citizenship education' and 'education for sustainable development' (ESD) have been reframed as 'skills' or 'competencies' (UNESCO, 2013; OECD, 2016a, 2016b), integrating them with the established OECD

paradigm that sees learning as a process of the competitive acquisition of skills (see the discussion of social and emotional learning (SEL) (WG2-ch8).

This approach now extends to 'creative thinking', due to be added to the PISA assessment from 2022 and defined as the 'competence' to engage productively in the generation, evaluation and improvement of ideas that can result in original and effective solutions, advances in knowledge and impactful expressions of imagination (Leksmono, Prihandoko and Murtikusuma, 2019). In this regard, the OECD approach focuses on two broad thematic content areas: 'creative expression' and 'knowledge creation and creative problem solving'. 'Creative expression' refers to instances where creative thinking is involved in communicating one's inner world to others through both written and visual forms, whereas 'knowledge creation and creative problem solving' involves the investigation of open questions or problems and the generation

... assessment today is both designed and implemented at various levels – from the classroom through school, locality, nation to the cross-national level.

of solutions that are original, innovative, effective and efficient (OECD, 2019)

## 9.2 .3

### IMPLEMENTATION OF ASSESSMENT

As the discussion of the OECD above underlines, assessment today is both designed and implemented at various levels – from the classroom through school, locality, nation to the cross-national level. Implementation at national and cross-national levels involves survey design as well as considerations of sampling, data collection and analysis. We examine further this cross-national dimension and its ideological and political dimensions later in this chapter, but it is important to bear in mind the influence of this global discourse as we consider the design and implementation of assessment at national and subnational levels.

Assessment at the individual and classroom levels can not only be implemented by the teacher but can also be accomplished by the student or their peers. Self-assessment (or assessment as learning) highlights the importance of self-monitoring processes during which students reflect on and evaluate their work or learning, judge the degree to which they reflect upon explicitly stated goals or criteria, identify strengths and weaknesses in their work, and revise accordingly (Andrade and Du, 2007). It is often suggested to be used in combination with other forms of assessment. For example, students can be required to present self-assessment in relation to previously agreed on criteria for activities such as class participation or presentation. Peer assessment is widely used in collaborative learning settings (Boud and Falchikov, 2007). It involves students providing feedback to other students on the quality of their work. The practice of peer feedback usually includes the assigning of a grade, or peer-related exchange and discussion



of student assignments (**Falchikov, 2013**). A significant amount of evidence suggests that students become more competent and gain confidence at peer and self-assessment practice (**Boud and Falchikov, 2007**). Moreover, **Taras (2010)** shows strong gains in examination scores when pupils were trained in both self and peer assessment.

The implementation of assessment is evolving rapidly with the advent of digital technology. With the aid of modern online

digital technology tools, including internet facilities, teleconferencing, videophones, multimedia systems and mobile technology, it is often claimed that teachers can now grade students' work quickly and easily, thereby saving more time for classroom instruction; but, in fact, pressures on teachers to be 'accountable' (and a proliferation of technologically enabled methods for enhancing 'accountability') frequently distract from engagement with students in the classroom (**Muller, 2018**). The latest technological tools notionally offer a quick

Artificial intelligence (AI) has also been hailed as promising enhanced possibilities for individualized learning and formative assessment.

assessment of students' work while providing instant feedback to students. Run by the OECD in 2018, the Teaching and Learning International Survey (TALIS) Video Study, through a direct video-recorded observation and lesson artefact collection, provides additional evidence on classroom teaching and instruction, and has been hailed as addressing the limitations of using teacher self-reported data (Ingram et al., 2020). However, the risks attached to reliance on such technology should also be clear: in some societies today, CTV cameras are becoming ubiquitous in university and school classrooms, where they serve the purpose of monitoring and controlling teachers and students. Likewise, smartphones and social media apps can be used to subject teachers to constant surveillance, either by an intrusive and controlling state, or by online mobs, or both (see WG2-ch6 on educational technology).

Artificial intelligence (AI) has also been hailed as promising enhanced possibilities for individualized learning and formative assessment

(UNESCO IITE, 2020). As a result of contemporary theoretical insights about teaching, learning and assessment (Gipps, 2002; Gibbs, 2006; Boud and Falchikov, 2007) as well as contemporary technological change, it is claimed that education will become more learner-centered and competence-based. But apart from the complex implications that introduction of new media may have for curriculum and pedagogy (see WG2-ch6, WG2-ch8), it is important to note that most developing countries are unable fully to utilize technology-enhanced approaches to assessment. Most schools in developing countries do not have a functional information communication and technology (ICT) unit; those that do are likely to be poorly equipped, while some may be equipped but are saddled with irregularities in the power supply. Furthermore, the use of digital tools for assessment can be problematic because of their reliance on decontextualized formats to assess learners' capacities with minimal room for observation of learning processes.



# 9.3

## Assessment and context: towards a global assessment regime?

Overall, there are two major approaches to understanding the relationship between assessment and context (**Unterhalter, 2016**). The first approach conceives of context as a linear system, in which assessment functions as a means of monitoring the input and output of education. In this

essentially mechanistic approach, assessment is mainly seen as a practice designed to enhance education's role in forming productive workers. The second approach considers context as a complex system where different aspects of sociocultural, economic, political and technological factors

Results of international large-scale assessments, as represented by PISA, have emerged as a reliable proxy for a nation's stock of human capital, and, therefore, future economic competitiveness.

intertwine, and assessment has the potential to provide a space for expressing aspects of what truly matters. This second approach acknowledges the important role of assessment in measuring outcomes and enhancing productivity, but also emphasizes its implications for rights, capabilities, equalities and relationships of power within and between societies. This latter approach is exemplified by efforts by Elaine Unterhalter and colleagues<sup>3</sup> in a project to develop an indicator framework to track aspects of gender equality in education in order to contribute to the measurement and evaluation of gender sensitive and inclusive learning environments. Through participatory discussions, the project has built on a range of research on gender, capabilities theory, education, poverty, intersectionality and inequality to develop innovative ways to evaluate gender equality in education, avoiding the limitation of relying solely on gender parity

measures. In this chapter, we also take the second approach to show how these multidimensional contextual factors shape and inform our understanding of the various forms of assessment, and their implications.

## 9.3 .1

### GLOBAL TRENDS IN ASSESSMENT

In the context of contemporary processes of globalization and the move towards a knowledge economy, the demand for improved comparative datasets in education has brought the authority of international reference frames to governance as part of what has been termed a 'comparative turn' (Martens 2007). Results of international large-scale assessments (ILSAs), as represented by PISA, have emerged as a reliable proxy for a

<sup>3</sup><https://www.genderddata.org/>

nation's stock of human capital, and, therefore, future economic competitiveness. Statistical sophistication has been paraded as a badge of scientific rigour, with claims that ILSAs incorporating modern psychometric techniques such as item response theory (IRT) make for improved comparability of results and depth of analysis across countries (Adams and Gonzalez, 1996). Multilevel modelling (Bryk and Raudenbush, 1992; Kreft et al., 1995) and structural equation modelling (Hayduk, 1987) are also portrayed as having enhanced the quality, depth and accuracy of results of a wide range of assessment methods, including cognitive diagnostic analysis. Such perceptions, however, often owe much to strategic narratives promoted by powerful institutions intent on securing and expanding their influence (Tikly, 2017; Ydesen and Grek, 2020; Auld and Morris, 2021).

Perhaps the most globally influential of these organizations is the OECD, which is why we devote special attention to it later in this chapter. But it is first pertinent to ask what features of

our contemporary international context have contributed to a situation in which it has fallen to 'an international economic policy organization' to 'lead the charge' towards global coordination of educational measurement – and what dangers this implies (Engel et al., 2019, p. 128). Crucial here are political and economic conditions and the ideological climate following the end of the Cold War, discussed in **WG2-ch1** of this 'Context' report. As the discrediting of state-based 'welfarism' paved the way to the ideological dominance of neoliberal, market-based approaches to economics and public policy i.e. new public management [NPM], educational opportunity rather than redistributive taxation or direct state support came to be portrayed as the ultimate guarantor of social justice. As discussed in **WG-ch1**, the concept of 'meritocracy', originally conceived as a dystopian fantasy, has been reframed as a utopian aspiration. If only all citizens could be offered the opportunity of a decent education, success or failure could be attributed



... positive reframing of meritocracy, coinciding with the massification of higher education and changing labour market trends in many societies, has hugely raised the political stakes surrounding educational assessment.

entirely to individual effort, and state welfare provision otherwise restricted to a minimal 'safety net'. Inequality could not only be justified 'meritocratically' but as a necessary spur to diligence on the part of learners and workers, and hence to economic productivity. Such, crudely put, is the essential logic of the new orthodoxy of which the OECD has become an influential exponent.

This positive reframing of meritocracy, coinciding with the massification of higher education and changing labour market trends in many societies (in turn linked to technological and economic developments discussed in **WG2-ch1**, **WG2-ch3**, **WG2-ch6**), has hugely raised the political stakes surrounding educational assessment. The belief that education holds the key both to maximizing national productivity and to distributing its benefits equitably has intensified pressure on governments to demonstrate that they can ensure delivery of a high quality of education to all their citizens. So too has the fact that, even as neoliberal

globalization eroded tariff barriers and promoted increasing cross-national regulatory alignment, education remained one area of public policy with implications for economic growth for which national states can be held entirely responsible (**Green 1997**). This does not necessarily imply direct state provision; neoliberal orthodoxy holds that the state's functions should be essentially regulatory, establishing and policing rules to enable a smoothly functioning market to serve 'consumers'. Since the 1990s, states have rarely divested themselves of existing public schools or colleges, but in societies where NPM has made the greatest inroads, management of these institutions has become increasingly marketized (see **WG2-ch3** on the political economy of education).

Just as the ideology of meritocracy sets an elevated premium on the measurement of learners' performance, so marketized management practices demand detailed measurement of the value of educational 'output' obtained in return for a given 'input'.

Like shareholders scrutinizing their company's annual report, taxpayers, parents and the politicians acting on their behalf (and, increasingly, private investors in the 'education industry', too) have been encouraged to demand 'accountability' from schools, teachers and educational officials, in the form of ostensibly transparent, scientific, quantifiable data.

## 9.3 .2

### TOWARDS A GLOBAL ASSESSMENT REGIME

These trends account for the huge appeal of the quantitative measurement of educational outputs offered by the OECD through its PISA tests, and for the introduction in many societies of increasingly elaborate systems of national testing. Across the English-speaking world especially,

the period since the 1980s (when the IEA introduced its Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS) has witnessed a drive to emulate what is seen as the test-driven success of East Asian systems of schooling.<sup>4</sup> Ironically, this has happened even while many East Asians themselves have voiced concerns about what they see as the excessive competitive intensity of their assessment regimes, and the effects of this – on children's stress levels, their physical and psychological health, household finances (as parents seek advantage in the race for credentials), socio-economic inequality and even decisions over fertility (**Vickers and Zeng 2017**). Legitimation for Western calls to 'look East' has come from claims for a correlation between performance in PISA tests and rates of economic growth (**Hanushek and Woessmann, 2008**), despite research demonstrating the evidential weakness of these claims

<sup>4</sup>Since the PISA tests began in 1997 (with the first results published in 2000), the work of the OECD in this area has increasingly overshadowed that of the IEA.



(Komatsu and Rappleye, 2017).

The early twenty-first century has thus witnessed a global intensification of assessment – not just of students and national systems of education, but also of teachers and schools (such as through PISA4U and PISA for schools) – driven fundamentally by an economic agenda that sees education as a vehicle for human capital formation. This agenda, strongly associated with the OECD, is in tension with the humanistic goals for education traditionally espoused by UNESCO. While not denying the vital economic importance

of education, in a series of major reports over its seventy-year history, UNESCO has elaborated a broader vision encompassing the intrinsic, as well as instrumental, relationship of learning to human fulfilment, the role of education in preparing students for active citizenship, and the promotion of tolerance and international understanding.

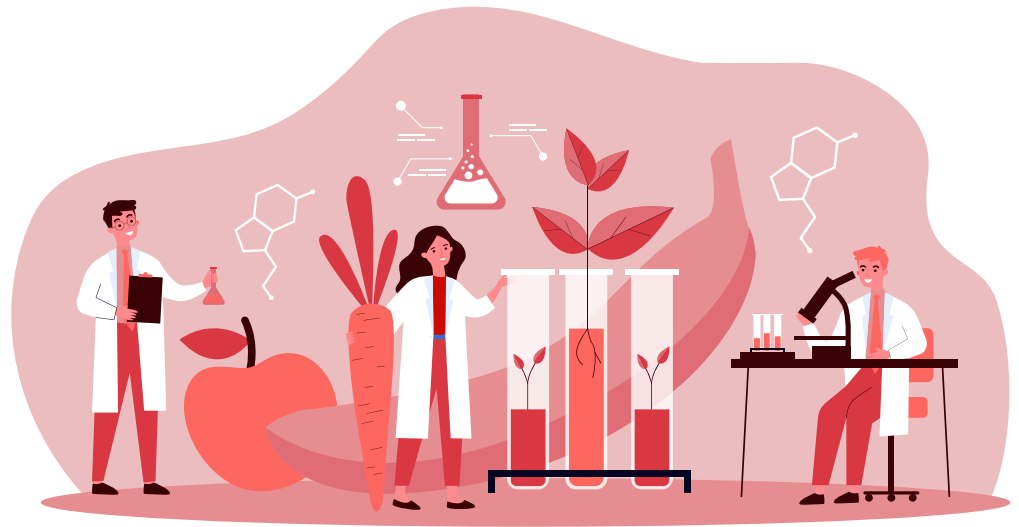
In recent years, however, there has been an increasing bending or enmeshing of these distinct economic and humanistic positions. On the one hand, in its educational work, UNESCO has come under growing pressure

The early twenty-first century has thus witnessed a global intensification of assessment – not just of students and national systems of education, but also of teachers and schools.

to demonstrate engagement with the hard-nosed world of statistical data, performance indicators and ostensibly ‘scientific’ analysis of educational inputs and outputs in the OECD mould (Elfert, 2018). Conversely, the OECD has responded to international calls, formalized in SDG 4.7, for education to promote ‘peace, sustainable development and global citizenship’, by devising new ways to define, quantify and measure ‘global competencies’ portrayed as promoting these goals (see WG2-ch8 for discussion of SEL in the context of curriculum). The relentless drive to quantify and measure, allied to the reframing of the socializing and humanistic aspects of learning as ‘skills’, enables talk of education’s role in promoting ‘sustainability’ and ‘global citizenship’ to be subsumed within a human capital formation discourse, effectively deflecting criticism of what remain its fundamentally economic premises (Auld and Morris, 2021).

Meanwhile, technological change, including AI and machine

learning, have influenced our understanding of various aspects of education and learning, including the potential for expanding and refining assessment techniques. Assessment is increasingly digitized, with the prospect of rendering it more adaptive to individual learning differences. So far, however, many experiments with online assessment have only changed the way of presenting the assignment by moving the questions from paper to screen and changing the assessors from teachers to machines (Wu et al., 2020). The implications of such ‘mechanization’ for the social or socializing functions of education, in the form of human interaction between teachers and students, and amongst students themselves, require careful consideration (Williamson, 2017). The urgent need for critical analysis of the potential and risks of a proliferating use of technology in assessment as well as teaching and learning has been heightened by the rapid expansion of online modes of educational delivery during the COVID-19 pandemic (see WG2-ch1, WG2-ch6, WG2-ch10).



# 9.4

## Place of science and evidence in assessment

The perceived validity of assessment procedures depends on claims to some sort of scientific and evidential basis, although definitions of what constitutes ‘science’ and ‘evidence’ can vary widely (Mislevy, 1994; Mislevy and Haertel, 2006; Hunsley and Mash, 2007; Pellegrino, Chudowsky and Glaser, 2001; Bennett, 2015). Taking a transformative perspective, the key question is how we can validate the outcomes of the assessment of aspects of child functioning. Therefore, implicit choices need to be made regarding what counts as ‘evidence’ based on claims that can be made from ‘science’. Over

recent decades, studies from the field of educational neuroscience (EN) have come to play an increasingly prominent role in public debate over assessment and education more generally around the world. It is therefore important to analyse not only the evidential basis for the claims advanced by neuroscientific research, but also the contextual factors that have influenced growing public interest in the application of neuroscience to educational assessment – and what this implies for assumptions about the nature and purpose of education.

Within the broader field of mind, brain and education, EN aims to use evidence from neural mechanisms of learning to optimize educational practices and policies.

However, conducting such an analysis is complicated by the fact, already noted in **WG2-ch7**, that social scientists and neuroscientists working on education operate with fundamentally different understandings of ‘context’. For neuroscientists ‘context’ appears to mean the set of external factors that impinge upon the experience of an individual learner, while for social scientists, analysis of ‘context’ implies examining the ways in which political, cultural, social and economic factors influence our educational institutions, practices and underlying beliefs. The extreme difficulty of reconciling these widely divergent interpretations of ‘context’ itself reflects the importance of ideology, politics and culture in shaping the disciplinary outlook of ‘scientists’ and ‘social scientists’. Following the exposition here of the relationship between context and assessment as seen from the perspective of neuroscientists, we return in the following section to an analysis of the ideological, political and cultural conditions that influence debate in this area.

## 9.4 .1

### IMPACT OF RESEARCH INTO EDUCATION AND LEARNING FROM A NATURAL SCIENCE PERSPECTIVE

Science and evidence have been hailed as foundational pillars of twenty-first-century education driven by the notion that empirical evidence is an ‘efficient indicator of knowledge and learning’ (**Wiseman, 2010**). EN, in particular, has emerged as a significant player in the field of science and evidence-based educational practices fuelled by the idea, as stated by Koizumi (**2004**), that education, aptly defined, is a ‘nurturing of the brain’ (**Howard-Jones, 2008**). Within the broader field of mind, brain and education (MBE), EN aims to use evidence from neural mechanisms of learning

EN supports the 'constructivist model' of education (typically associated with the 'formative' approach to assessment) where students should be engaged actively in the learning process and have an agency to guide their learning.

to optimize educational practices and policies. It has been claimed that the possibilities afforded by neuroscience in learning offer great opportunities, although advocates stress that its applications should be specific and supported by well-controlled experimental data that can be translated to classroom environments with success (**Wolfe and Brandt, 1998; Geake, 2008**). EN has been portrayed as supporting the 'constructivist model' of education (typically associated with the 'formative' approach to assessment) where students should be engaged actively in the learning process and have an agency to guide their learning (**Caine and Caine, 1991; Sylwester, 1995; Jensen, 2005**). EN has provided biological support to several psychological theories on learning but it is sometimes argued that the lack of specificity makes it difficult to translate such biological findings into classroom settings (**Bruer, 1997, 1999, 2016**). Direct solutions to learning-related real-world problems, faced in the classroom, have been lacking and it is likely that neuroscience alone cannot provide such direct solutions.

Furthermore, there are known methodological issues in the field of cognitive neuroscience that can extend to EN if proper vigilance to confirmation bias is not undertaken at policy levels, that is, focus should be on assessing results on the basis of sample and/or effect sizes and be wary of over promises based on correlation results and low replication rates (**Yarkoni, 2009; Poldrack, 2012; Barch and Yarkoni, 2013; Button et al., 2013; Szucs and Ioannidis, 2017, 2020; Turner et al., 2018; Huber, Potter and Huszar, 2019**). Before adopting any brain-related evidence at the policy level, it is important to ensure that EN is able to address a specific problem; it should also be tested in large-scale student populations, and replicated.

In an early example of a successful EN application, functional neuroimaging of younger students was found to require increased working memory and attentional resources to perform mathematical operations at the same accuracy levels as older students (**Rivera et al., 2005**). The information that younger and older students differ



in the cognitive resources used to perform arithmetic could, in turn, help teachers in the classroom, to design different teaching strategies for students based on their age. In the above example, neuroscience findings (e.g. brain activations) were used to study behaviour (mathematical problem-solving), establishing an indirect link between neuroscience and classroom learning strategies. However, direct links between neuroscience and learning can also be studied, for example, the impact of different variables like metabolism, exercise, nutrition, stress hormones or environmental pollutants on brain function in areas important in learning can be studied without behavioural assessment (Sigman et al., 2014; Feiler and Stabio, 2018; Thomas, Ansari and Knowland, 2019). In a similar vein, neuromodulation can be used to assess several contextual factors in its usage, for example, inequalities in education (Knowland, 2020). It is important to note here, although the focus of EN is primarily on the bidirectional relationship between brain changes and learning and education, biology

is merely the substrate on which social and cultural contexts act to cause changes (see WG3-ch5 for a discussion on ‘Foundational language skills for literacy’). For policy decisions and practices stemming from EN to be successful, integration of sociocultural context is of great importance; for what we learn from EN studies in Western, Educated, Industrialized, Rich and Democratic (WEIRD) (Henrich, Heine and Norenzayan, 2010) nations cannot be successfully implemented in the rest of the world without context-dependent adaptations (see WG2-ch7 on critical challenges in EN).

## 9.4 .2

### SCIENCE-BASED RESEARCH AND FORMATIVE ASSESSMENT

Naturally, at all education levels, understanding students’ learning

In a similar vein, neuromodulation can be used to assess several contextual factors in its usage, for example, inequalities in education.



Considering the importance of formative assessments in student learning and academic progress, designing optimal assessment programmes is a key concern.

profiles and their repertoire of learning strategies is of great importance. It allows the teacher to remain informed about the students' baseline, progress and process of learning. Formative assessment programmes seek to allow teachers to appraise exactly this – how well the student is performing on a path to reach intended goals. This should be distinguished from a formative assessment that uses judgements about how to improve programme effects (OECD, 2005). Considering the importance of formative assessments in student learning and academic progress, designing optimal assessment programmes is a key concern. Behavioural, psychological and neural data from EN studies can provide a vital understanding of the mechanisms underlying learning strategies and contribute to the design of successful formative assessment programmes. Learning begins as an amalgamation of several interactive and multidimensional cognitive processes that are essential for academic as well as lifelong learning, including working memory, motivation,

reward, selective attention, visual and auditory processes, executive functions, emotion regulation, awareness and reflection (Bunge and Souza, 2009; Price, 2012; Menon, 2015; Florensa, Duan and Abbeel, 2017). Neuroscientists stress that understanding the underlying basis of the interaction between learning and context can help educators and teachers guide the contextual learning environment to affect biological changes towards intended learning outcomes. However, it is not straightforward to translate neuroscience findings into instructional practices that improve learning outcomes (Bowers, 2016a), and for the field of EN to demonstrate a meaningful contribution to improving learning would require concerted efforts between different stakeholders (Smeyers, 2016; Thomas, Ansari and Knowland, 2019).

Although it is claimed that the future of the EN field opens up several promising avenues to inform teaching practices and, in some instances, redefine educational frameworks (Shore

... insights from neuroscience are used to make formative assessment tools more adaptive to identify both the differences in students' learning processes and the necessary instructions required to bridge any gap in such processes.

and Bryant, 2011; Smedt, 2018), for EN findings to deliver on their promise and be translated to real-world large-scale classroom practices, lab-based experiments need to be adapted accordingly and replicated rigorously (Seghier, Fahim and Habak, 2019). A related major finding from EN is the evidence of 'individual differences' in brain-behaviour processes in learning underlying the complex issue of generalizing the applicability of EN findings over large groups (Posner and Rothbart, 2009). In past decades, the OECD has been interested in the application of findings from EN. At an individual level, insights from neuroscience are used to make formative assessment tools more adaptive to identify both the differences in students' learning processes and the necessary instructions required to bridge any gap in such processes (OECD, 2006; WG3-ch3). It is argued that neuroscience methods can be applied to individual learners to grasp their depth of learning and emotion regulation capabilities, and with intelligent designing, such individual focus

can provide powerful diagnostic tools for formative assessment and personalized learning (OECD, 2007, p. 156). Indeed, there is neuroscientific evidence that a large number of brain networks function in an overlapping and dynamic manner to bring about learning and skill acquisition in different domains of knowledge, ranging from academic disciplines to social and emotional learning (SEL) (Thomas, Ansari and Knowland, 2019; Dahl, Wilson-Mendenhall and Davidson, 2020). However, it should be noted that more research is needed to establish whether such findings can ultimately validate the claims the OECD has made as regards the potential of EN in educational assessment (see WG2-ch7 and WG3-ch3 for discussion on 'Why neuroscience is relevant to education').

Learning can generally be categorized according to its nature, perceptibility and level of depth (Rogiers, Merchie and Keer, 2019). As the learner becomes more advanced, the depth of previously learned skills influences the ongoing learning process at

...literacy can be achieved through different developmental pathways and across the lifespan, thus helping in building different learning assessment tools and teaching instructions for children, adolescents and adults.

both the cognitive and neural levels (**Hinton, Miyamoto and Della-Chiesa, 2008**). Repeated testing and delayed feedback has been shown to increase long-term memory retention of newly studied text both in lab-based and classroom environments (**Dunlosky and Nelson, 1992; Roediger and Karpicke, 2006a, 2006b**) underlying the importance of continuous formative assessment in student learning, recall and overall academic achievement (**Ozan and Kincal, 2018**). Educators have also advocated for dynamic testing (feedback while the test is being conducted) as a better means for assessing students' strengths, weaknesses and learning potential in different cognitive domains when compared to static testing (**Grigorenko and Sternberg, 1998; Sternberg and Grigorenko, 2001, 2002**). This makes dynamic testing and formative assessment (an assessment method that embodies the idea of continuous as well as dynamic testing) one of the testing methods that promotes learning and not just test learning. EN has also helped in rejecting the neuromyth of 'critical periods' of learning

with the more appropriate 'sensitive periods' in learning (**OECD, 2007, p. 122**). This has led to the understanding that literacy can be achieved through different developmental pathways and across the lifespan, thus helping build different learning assessment tools and teaching instructions for children, adolescents and adults (**Frith et al., 2011; Parisi et al., 2019; also see Meltzoff et al., 2009; Varma, McCandliss and Schwartz, 2008; WG2-ch7 and WG3-ch3 for further reading on EN**).

## 9.4 .3

### SCIENCE-BASED RESEARCH AND SUMMATIVE ASSESSMENT

Summative assessment assesses student learning at different stages of the education timeline and provides for a learner's accountability. Summative assessment mostly uses

NAPLAN and similar standardized tests, which are highly consequential, have faced significant criticism due to their aim to assess student learning in a very short time but with results that have significant ramifications.

standardized tests to measure student learning, and high-stakes examinations used as summative assessment tools might have life-changing consequences on a learner's education profile going forward. For example, NAPLAN scores in Australia, a score derived from summative assessment, determine federal funding for schools, exert a significant effect on learners' future educational opportunities and might also be linked to teacher pay (Ford, 2013; Harris et al., 2013; Smith, Parr and Muhidin, 2019). NAPLAN and similar standardized tests, which are highly consequential, have faced significant criticism due to their aim to assess student learning in a very short time but with results that have significant ramifications. The current discussion on summative assessments raises questions of how, where and what is being assessed. And if standardized tests are true predictors of human flourishing. Innovative research in brain sciences might provide answers to some of these questions. Research from EN and social

sciences show that individual differences in biology, as well as sociocultural factors, play a key role in determining different levels of learning, that is, there are multiple ways of knowing (Thomas, Ansari and Knowland, 2019; Dahl, Wilson-Mendenhall and Davidson, 2020; WG3-ch2, WG3-ch3). This raises the question whether standardized tests can completely assess the learning abilities and outcomes of a diverse student population. Additionally, the biological response to stress while undertaking high-stakes anxiety-inducing assessment might play a major role in determining assessment scores.

The OECD introduced collaborative problem solving (CPS) in the assessment protocol of its PISA tests in 2015 (Mo, 2017) on the basis that CPS had been shown to help develop effective pedagogical strategies and learning technologies (Law and Tsang, 2019, p. 165). The PISA framework assesses two dimensions, cognitive and collaborative processes, required for CPS. In the PISA CPS assessment, each student

Children are the heir to the social, moral and ethical responsibilities of the global future.

interacts with a computer agent to complete the assessment. In the collaborative setting of CPS, the social and cognitive processes are assessed as different domains required for achieving CPS successfully. However, it is unlikely that at the neural level these two domains are activated independently of each other and more likely that social and cognitive brain networks interact to bring about CPS. Findings from developmental neuroscience can inform the trajectories of development in the social and cognitive processes in the developing brain (WG3-ch2).

## 9.4 .4

### TOWARDS A SCIENTIFIC WINDOW ON CHILD FLOURISHING

Children are the heir to the social, moral and ethical responsibilities of the global future. To fulfill the

responsibilities of their future role, not just as human capital in economic developments, but as gatekeepers of the planet's health and well-being, there is growing consensus that childhood education needs to be grounded in principles that lead to human flourishing. The current generation of children faces several critical issues globally that threaten their future health and flourishing, including, but not limited to 'Climate change, ecological degradation, migrating populations, conflict, pervasive inequalities, and predatory commercial practices' (Clark et al., 2020). One of the aims of future educational frameworks is to create learning environments that promote childhood flourishing that can build towards a resilient future generation capable of countering and adapting to these growing threats and fulfill their shared global responsibilities. The context of the evolved developmental niche plays a pivotal role in the successful learning of SEL. A classroom that is positive, congenial, empathic and stimulating is a necessity for



achieving these skills. Conversely, a stressful environment can result in the activation of the biological stress cycle that can eventually lead to reduced social-emotional and academic performance. Negative emotions such as anxiety and stress, often associated with language acquisition or mathematics learning, can be regulated through programmes focusing on SEL (**Dresser, 2012; WG3-ch4**).

A universal educational framework focused on childhood flourishing must not only be limited to academic knowledge acquisition and performance achievements but constitute learning of adaptive and positive social and emotional responses that are prerequisite to childhood well-being and good academic

performance. Several developing brain networks are implicated in the cognitive processing of the different domains of SEL (**Dahl Wilson-Mendenhall and Davidson, 2020**). Childhood flourishing is achieved when a child can realize their full potential related to cognition, social and emotional interaction, and linguistic and motor skills (**WHO, 2018, p. 83**). In the classroom environment, the focus should be on building and maintaining social relationships, recognizing emotions in self and others, regulating strong emotions, making responsible decisions, having self-agency, developing effective and collaborative problem-solving skills and making empathic responses (**Committee for Children, 2019**). Both proximal forces like maternal health and early caregiver–child relationships,

... it has been observed that cultural traits that allow human flourishing co-evolve with biology and shape biological affordances.

and, distal forces including culture, politics, ideology, pandemics, economics and, increasingly, the climate (Watts et al., 2019; Black, Lutter and Trude, 2020) contribute significantly to the development of either adaptive or maladaptive social-emotional responses at the neurological and behavioural levels (Gleason and Narvaez, 2019).

Considering culture as a context in which SEL takes place, it has been observed that cultural traits that allow human flourishing co-evolve with biology and shape biological affordances (Aggarwal, 2013). For example, basic emotions and the social construction of emotions are represented by the dual processes of genetic and cultural inheritances (Chiao, 2015). Studies in cultural neuroscience have used transcultural neuroimaging (Han and Northoff, 2008) to show that there is a 'looping effect' in the dynamic interplay between culture and biology (Vogele and Roepstroff, 2009). Using cultural neuroscience as a tool to understand how emotion recognition varies across cultures, Chiao and colleagues

(2008) showed that the amygdala, the area primarily involved in the processing of emotions and memories associated with fear, preferentially reacts to culturally congruent fearful faces compared to culturally incongruent fearful faces. Systematic and effective assessment procedures, as implemented by formative assessment, should be carried out regularly for monitoring and evaluation of the successful implementation of educational programmes with a focus on child flourishing (Ferreira, Martinsone and Talic, 2020).

## 9.4 .5

### SCIENCE AND EVIDENCE-BASED POLICY AND PRACTICE

As noted above, an intensified focus on the role of science and evidence in education was rooted in concerns over the ability of

However, the rigour of statistics and evidence of ‘what is impactful’ has resulted in exclusion criteria that reject impactful interventions where such statistical significance is difficult to demonstrate.

education in preparing future citizens who can maximize financial outcomes based on HCT and address inequities in quality education (**Bennet and Gitomer, 2009**). In the United States (USA), the No Child Left Behind Act (NCLB Act) of 2001 was one of the first policies that prioritized the role of randomized studies and ‘scientific evidence’ in data-based policy decision-making. In the next two decades, the relevance of science and evidence in education policies (particularly, in middle- and high-income societies) has been growing steadily (**Ross and Morrison, 2020**). Results from the OECD’s PISA (**OECD, 2006**), which reported below-average performance of US students in comparison to other industrialized nations with whom the USA competes at an economic level (**Lemke et al., 2004**), was a major driving force for the emergence of evidence-based policy decisions in the USA. The 2015 Every Student Succeeds Act focused on raising the standards of research-derived scientific evidence required to evaluate the education programmes of schools. As Ross and Morrison

(**2020**) point out, this focus on science and evidence on evaluation has resulted in an evidence-based back-up of assessment tools, such as consumer access to evidence databases and supporting intervention programmes that are backed by evidence for funding. However, the rigour of statistics and evidence of ‘what is impactful’ has resulted in exclusion criteria that reject impactful interventions where such statistical significance is difficult to demonstrate (**Asen et al., 2013**). For example, growth in academic performance following goal-directed interventions is easier to achieve in terms of statistical significance than intervention programmes that focus on SEL, although the importance of the latter in lifelong success is well documented (**Zins and Elias, 2007; Morrison, Ross and Reilly, 2019**). (Also see Farley-Ripple et al. [**2018**], on how scientific evidence might be used in policy decision-making). In their case study report on the ‘technology diffusion’ initiative, Ross and Morrison (**2020**) also observe the changing attitudes of stakeholders across the cross-section of the initiative timeline.



... policy decisions that incentivize teachers solely based on academic testing scores can deny funding and undertaking of classroom research that do not follow the ambitions of student test scores.

The increase in digital comfort and skill development, though initially appreciated, became less relevant for stakeholders over time; the primary focus was pivoted towards the programme's efficacy concerning educational achievement, raising questions about the intended outcomes of educational initiatives beyond academic performance.

Specific to the role of neuroscience in education policy decision-making, Shore and Bryant (2011) advise educational policy-makers to focus on policies that can redefine classroom practices that might limit the use of neuroscience findings. One of the significant findings from biology has shown the importance of positive, congenial and empathetic learning environments in learning and academic performance (Kort, Reilly and Picard, 2001; Jensen, 2005; Sousa, 2006). It was observed that a stern and/or low-quality learning environment results in increased stress levels and consequently higher levels of the stress hormones adrenaline and cortisol. This increase in stress hormones acts on specific brain networks, resulting in negative

emotions and reduced EF, both of which significantly impede effective learning and academic performances (see WG2-ch5 for further discussion on the biological response to stress and its effect on learning). Along with the findings of social and emotional contexts that act on the brain to mediate learning (Dirkx, 2008; Hinton, Miyamoto and Della-Chiesa, 2008; King and Chen, 2019; WG2-ch4), these findings have provided meaningful evidence on how educational policy and practices must be shaped to create positive learning environments. However, Geake (2008) reminds us that neuroscience findings in lab-based experiments by themselves might not be relevant to education and teaching practices; instead real-world implementation providing 'a stronger rationale as to why particular styles of teaching and certain strategies are more effective in reading and writing, for example than others' is required. Jalongo (2008) also discusses the role that economics and politics play in prioritizing one set of findings over another. Additionally, policy decisions that incentivize teachers solely based on academic testing scores can deny funding and

## Education policy-makers must consider conducting studies and adapting study designs in reference to the cultural context to design successful assessments.

undertaking of classroom research that do not follow the ambitions of student test scores (**Farstrup and Samuels, 2002**). This same complex dynamic between scientific evidence and politics also comes into play in sanctioning or subverting the use of neuroscience findings in educational policy. The decision is often based on individual biases within the political system and not on the robustness, technical merit or utility of the conducted research (**Britto, Cerezo and Ogbunugafor, 2008**).

The growing role of neuroscience in policy decision-making has not been limited to the USA. Several initiatives across different nations, for example, the United Kingdom (UK), France, Denmark, Australia and Singapore, have been implemented to integrate neuroscience with policy across different fields including military, law and education (**Pykett, 2015**). However, beyond the developed nations, world-governing bodies like the UN and the World Bank have also invested in developing a framework for policy programmes influenced by neuroscientific evidence. From this perspective,

Hinton, Miyamoto and Della-Chiesa (**2008**) have laid down a few important points for brain-informed education policy implications for consideration by education policy-makers including a) building towards rich learning environments, b) embedding guidelines for formative assessment into the curriculum to help spur ability growth, c) considering the interplay between emotion and learning, d) considering sensitive periods for language learning and, e) considering neuroscience findings to inform reading and mathematics instruction. However, the role of cultural context is often ignored when considering educational policy implications. EN research still has a great deal to accomplish when it comes to comparing and contrasting findings across different cultures. Findings from one culture might not be transferable to other cultures (**Zhou and Fischer, 2013**). Education policy-makers must consider conducting studies and adapting study designs in reference to the cultural context to design successful assessments.



# 9.5

## Assessment and ideology

The preceding section reflects  
views of the relationship

between context and educational  
assessment from the perspective of



... assessment is not just a value-neutral instrument for promoting and calibrating effective learning; it also functions as a tool of political and social control, in ways that go unrecognized in much public debate over education.

the laboratory-based scientist. But whereas scientists or engineers may approach contextual factors as so much grit or oil in the assessment machinery, obstructing or facilitating its efficient operation, sociologists, historians or philosophers are typically inclined to ask: who has designed that machinery, with what purposes, in whose interests, and with what effects? As numerous citations in the previous section illustrate, the OECD has played a leading role in promoting the potential of EN, despite the fact that many of the claims made for the significance of neuroscientific research rest on faith in future 'promise', rather than on already existing evidence of its capacity to transform the assessment of learning. The OECD appears animated by the hope that neuroscience can supply an armoury of scientifically based interventions calculated to maximize the efficiency of the learning process (and hence the productivity and adaptability of future worker-citizens worldwide), but that hope itself can be considered a reflection of a profoundly individualistic, human

capital-oriented conception of the purposes of education.

Assessment, in other words, cannot be understood apart from its ideological, political and cultural context. As noted at the beginning of this chapter, assessment is not just a value-neutral instrument for promoting and calibrating effective learning; it also functions as a tool of political and social control, in ways that go unrecognized in much public debate over education. And to the extent that scientific and technological advances expand and refine the range of assessment techniques available to state and corporate actors, this can be (and, in some societies, is already proving) not liberating and empowering, but profoundly detrimental to the capacity of citizens to live dignified, fulfilling lives. How new approaches to assessment are deployed, and with what effects, has far more to do with politics, culture and vested interests than with purportedly objective 'science and evidence'.

## BOX 1

A notable example of how policy-makers can be lured (or bamboozled) by scientific claims is the fiasco that occurred in the UK in the summer of 2020 when, having cancelled regular public examinations because of COVID-19, the government decided to assess students on the basis of teachers' predicted grades adjusted by algorithms. Ministers had apparently assumed that 'algorithms' were, by definition, scientific and objective, without understanding that an algorithm is only as sound as the assumptions made by the programmers who designed it. When results were released in July that year, they were greeted by a wave of protest as students, parents and teachers pointed to numerous examples of seemingly arbitrary re-grading, and the government was forced to backtrack, abandoning the use of the algorithm.

## 9.5 .1

## SOCIOCULTURAL AND POLITICAL ARGUMENTS

'Culture', like 'context', is a concept subject to widely differing interpretations by researchers of different disciplinary backgrounds (or, indeed, disciplinary cultures). In much of the literature concerning assessment emanating from the OECD, reflecting the statisticians' 'thirst for reliable metrics', culture tends to feature as one among a number of 'factors' to be taken into account in ensuring the universal applicability of an assessment regime. But an historian, anthropologist or comparative social scientist is more likely to see culture not merely as a factor or set of factors influencing how effectively students learn, but as an ethical and philosophical framework shaping assumptions about what is worth learning in the first place (see WG2-ch8 for discussions on curriculum and pedagogy) (Alexander, 2000).

An approach to culture that aligns with or reflects some of the assumptions informing ambitious, cross-national projects of educational assessment is 'world culture theory'. This theory posits that as modernity, and modern education systems, proliferate around the world, educational ideas, practices and institutions are increasingly converging upon a single global model. As evidence, world culture theorists frequently point to the similarity of school timetables and curricular categories in different countries (**Meyer, Kamens and Benavot, 1992**). However, critics have pointed out that, once one opens school textbooks or enters classrooms in different countries, the ways in which ostensibly similar terms or concepts are interpreted in practice widely differ (**Carney, Rappleye and Silova, 2012**). There is, in fact, no convincing evidence of worldwide convergence in our culturally-informed assumptions about what education is, or should be, for, and how teaching and learning should be conducted.

At the same time, there are dangers in overstating the extent or immutability of cultural differences. Cultures are not immutable essences attached to eternally fixed human communities, but evolve in complex relationships with the political, cultural and socio-economic forces at work within any community or society, and beyond it, in its interactions with the wider world. The idea that any particular community possesses an entirely unique and incommensurable cultural identity that precludes meaningful comparison with, or borrowing from, other communities is a fallacy frequently invoked by authoritarian rulers to delegitimize and stifle dissent. The Indian thinker Amartya Sen has dubbed this 'the Lee thesis', after the late Singaporean elder statesman Lee Kuan Yew, who was fond of invoking the nebulous concept of 'Asian values' to justify his idiosyncratic interpretation of democracy and civil liberties (**Sen, 1999**). Culture, then, is vitally important, but its importance needs to be understood

Cultural sensitivity in educational debate is important...to ensure respect for, and enhancement of, the agency of underprivileged and disempowered communities.

in historical and political contexts. Cultural sensitivity in educational debate is important not in order to avoid offending autocrats, but to ensure respect for, and enhancement of, the agency of underprivileged and disempowered communities. As the discussion below of Africa and PISA in low- and middle-income countries illustrates, this has emerged as a key challenge for ILSAs.

A number of researchers have invoked the concept of ‘indigenous education’ to signify their recognition of how ILSAs such as PISA struggle to encompass or account for the cultural differences that underlie the enormous variation in educational beliefs, practices and institutions (e.g. Gohl, Gohl and Wolf, 2009; Brock-Utne 2016). Indigenous education and assessment align with twenty-first century skills by involving teachers and students as co-constructors of education and valuing the interconnectedness of content and context (Munroe and Toney, 2013). Therefore, the inclusion of indigenous education

and assessment procedures requires acknowledgement of the existence of a multiplicity of forms of knowledge rather than a particular ‘standard’ benchmark system often put forward by the West. For example, cultural and social norms affect how test-takers comprehend and interpret the wording of the questions contained in the tests. And correspondingly, how the learners make sense of the test items can be influenced by the values, beliefs, experiences, communication patterns, teaching and learning styles, and epistemologies of the cultural values inherent in their societies (Solano-Flores and Nelson-Barber, 2001).

9.5 .2

## ASSESSMENT AND HUMAN CAPITAL

As noted in previous chapters (WG2-chs1--3), discussions of assessment results often implicitly

ILSAs such as PISA struggle to encompass or account for the cultural differences that underlie the enormous variation in educational beliefs, practices and institutions. Indigenous education and assessment align with twenty-first century skills by involving teachers and students as co-constructors of education and valuing the interconnectedness of content and context.

or explicitly frame education as an investment in human capital to promote economic development. Conventional HCT asserts that the education level of a country is an important factor for explaining national economic growth (Schultz, 1961, 1971; Becker, 1964). Unlike traditional factors (such as labour and land), human capital – pertaining to the knowledge and skills workers have acquired as a result of education – contributes to productivity and thereby to earnings at an individual and collective level. The increase in human capital stock thus leads to economic growth.

While early estimates have mainly used school attainment (or years of schooling) as measures of human capital (e.g. Mincer, 1974; Psacharopoulos, 1994; Psacharopoulos and Patrinos, 2004), more recent studies have turned the focus to cognitive skills and to the use of ILSA scores to arrive at more accurate measures of human capital. In particular, economists Eric Hanushek and Ludger Woessmann have introduced PISA test-score measures into

growth regressions, arguing that the achievement measure is substantially more positively associated with economic growth than the attainment measure.

*For instance, [a]fter controlling for the initial level of GDP per capita and for years of schooling, the test-score measure features a statistically significant effect on the growth of real GDP per capita in 1960–2000. According to this specification, test scores that are larger by one standard deviation (measured at the student level across all OECD countries in PISA) are associated with an average annual growth rate in GDP per capita that is two percentage points higher over the whole forty-year period. (Hanushek and Woessmann, 2008, p. 638)*

For a long time, international development agencies have pursued the expansion of schooling as a primary component of development (Mundy, 1998, 2006). The evidence that Hanushek and Woessmann (2008) provide seems to speak to the significance



of cognitive skills and of improvement and measurement of learning outcomes. This goes a long way toward explaining why key agencies, such as the World Bank and the OECD, in addition to extending their influence and power (see further discussions below), have come to prioritize the use of ILSAs to gauge learning and drive development in low- and middle-income countries.

While this cognitive-economic model based on international educational performance data has been much debated and critiqued in the academic literature (e.g. **Komatsu and Rappleye, 2017, 2019; Feniger and Atia, 2019; Patel and Sandefur, 2020**), it continues to gain momentum. Klees (2016, p. 644) comments that it ‘has been ubiquitous and widely accepted as an important mechanism for educational planning, evaluation, and policy making’. The result has been that ‘earlier uncertainties about how education works, how it impacts society, and how to best allocate scarce resources are being quickly replaced by contemporary certainties that raising test scores

will result in higher levels of economic growth (GDP)’ (**Komatsu and Rappleye, 2017, p. 1**).

## 9.5 .3

### GLOBAL ASSESSMENT AND IDEOLOGY

As **WG2-ch2** has emphasized, a key challenge for educators and education policy-makers today is the absence of an ethical and political vision of education that supports a transition towards a more equitable and sustainable world (see **Beech and Rizvi, 2017**). The emphases or biases of our assessment regimes are a key part of this problem. A dominant conception of education that sees it above all as an instrument for maximizing human capital formation is associated with an approach to assessment conceptualized as the summative evaluation of knowledge and skills above all in the fields of literacy, mathematics and science. Assessment of education at

national and cross-national levels has increasingly been reduced to children's profiles of scores on standardized tests in these areas of knowledge. While these subjects are considered most relevant to ensuring readiness for employment, knowledge in these areas is also relatively susceptible to quantitative measurement, and relatively (though far from entirely) comparable across systemic and cultural boundaries.

However, the focus of political and media attention on the results of tests of these more readily 'testable' knowledge domains can lead to neglect and marginalization of other subjects that matter, such as the arts, humanities, physical education and social skills. Even more important is the notion that assessment should not involve the evaluation of a series of disciplinary subjects, but the human subject itself. Therefore, there is the need to focus not only on cognitive targets aiming at what students learn but also on ontological targets aiming at what students become (Dewey, 1923). In this respect, a

distinction can be made between mimetic and transformative educational traditions (Jackson, 1986). Mimetic traditions relate to the transmission of factual and procedural knowledge, whereas transformative traditions relate to the transformation of one kind to another in the person being taught. In future assessment protocols, it is important that these two traditions are reconciled to contribute to the social transformation of the next generation of children.

Insight into ideological aspects of assessment in education at a global level can be better understood from local case studies. Therefore, we focus on assessment and ideology in two major parts of the world: China and Africa. China's long history of examinations has already been noted above, and high-stakes public examinations, notably the National College Entrance Exam (NCEE, known as Gaokao), play a crucial role in underpinning perceptions of justice in the distribution of social opportunity. The recent reform of Gaokao has attempted to offer

India withdrew from PISA in 2009 after performing poorly and claimed that the test had not been sufficiently contextualized.

the rights to choose examination subjects for students in order to promote their personalized development. However, it should be noted that such reforms reflect how examinations in China may serve as sites of struggle against widespread sociocultural beliefs, as the communist regime seeks to use assessment to reinforce and legitimate its authority. Recent work on the politics of assessment in China characterizes the regime as an ‘Assessment State’ which uses increasingly sophisticated techniques of monitoring, surveillance and assessment of adults, as well as children, to accomplish hierarchical reordering of society (Wan and Vickers, 2021). It differs from Africa, where most local authorities are endeavouring to liberate themselves from the shackles of their colonial masters (Unterhalter, 2009). For example, Brock-Utne (2016, p. 40) notes that educational researchers ‘have been constantly debating what quality in education may mean in their own context and how it should be assessed’. She suggests that, in order for Africa to build up an education system that

adheres to their own values (e.g. care for others and cooperation), ‘it is critical both to teach in the languages learners speak and understand and to avoid the Western testing regime’ (2016, p. 41). It is suggested that the world intellectual community and major development agencies accord some forms of recognition to African Indigenous education and assessment (Obanya, 2007, p. 25).

## 9.5 .4

### THE OECD AND THE EMERGENCE OF A GLOBAL ASSESSMENT REGIME

Since PISA was first conducted in 1997, it has received considerable media coverage and attention from politicians and policy-makers in many nations. And as it continues to expand, the OECD has become a recognized global provider of technical expertise in the measurement of schooling

...many scholars have argued that the challenges that low- and middle-income countries faced in engaging with PISA have not been resolved by PISA-D.

performance in both member and non-member nations. With the introduction by UNESCO of the global education agenda in 2015, which identified minimum standards of quality for all countries to be delivered by 2030, the OECD has sought to promote PISA as the universal learning metric. The intent of the OECD is to have 170 nations participate in PISA by 2030, and the means to achieve this is the programme called PISA for Development (PISA-D). Whilst the OECD has enrolled a few low- and middle-income countries to join PISA in the past decade, it also faces significant challenges in these contexts (Carr-Hill, 2015; Lockheed, Prokic-Breuer and Shadrova, 2015). For example, India withdrew from PISA in 2009 after performing poorly and claimed that the test had not been sufficiently contextualized (see Edwards, 2019). Thus, the OECD, in late 2013, introduced PISA-D to make PISA more accessible and relevant by extending the PISA test instruments to ensure wider coverage in performance levels, developing contextual

questionnaires to effectively capture diverse conditions, and establishing approaches to include out-of-school youth in the assessments (Adams and Cresswell, 2016). Whilst the initiative was portrayed as helping to identify how PISA could best support evidence-based policy-making in developing societies and contribute to the monitoring of SDG 4 targets and indicators, specifically those related to learning outcomes, many scholars have argued that the challenges that low- and middle-income countries faced in engaging with PISA have not been resolved by PISA-D (see Brock-Utne, 2016; Kaess, 2018; Auld, Li and Morris, 2020; Li, 2021; Rutkowski and Rutkowski, 2021). Moreover, it is also too early to judge whether PISA-D findings will contribute to reforms aimed at improving the quality of learning in the participating countries.

In addition to the above, PISA for Schools has been introduced as an attempt to extend the relevance of PISA and the reach of the OECD into local school practices (Lewis, 2017). This alternative

... participation in PISA by developing countries is driven by the motive to be put on the global education map, a decision primarily driven by political reasoning.

framework assesses not only school performance in reading, mathematics and science against international schooling systems, but also promotes examples of best practice from world-class schooling systems and, in turn, the policy expertise of the OECD itself.

The global expansion of PISA raises several questions about the influence of ILSAs on global policy and practice in education (**Addey and Sellar, 2018**). Often, participation in PISA by developing countries is driven by the motive to be put on the global education map, a decision primarily driven by political reasoning (**Grek, 2009; Kamens, 2014; Wiseman and Chase-Mayoral, 2014; Addey, 2015**), but also for reasons of accountability, transparency and social development leading to ‘governance by comparison’ (**Novoa and Yariv-Mashal, 2003; Grek, 2009; Addey and Sellar, 2018**). Addey and Sellar (2018) summarize, through data analyses, the seven factors that drive participation in larger-scale assessments by nations: (1) evidence for policy;

(2) technical capacity building; (3) funding and aid; (4) international relations; (5) national politics; (6) economic rationales; and (7) curriculum and pedagogy. These demonstrate the complex dynamics between ILSAs and epistemological and infrastructural global governance. The World Bank, along with associated grant agencies and governing bodies, is a major advocate of large-scale assessments in developing countries, often projecting assessment scores as evidence-based achievement markers for what works (or not) in education. However, interpretations of PISA results are ambiguous, and variables at national levels lead to what are known as ‘multiple truths’. Furthermore, there are several other concerns regarding the use and abuse of such global summative assessments (**Lockheed, 2013**). These concerns include the ‘floor effect’ of PISA, where the assessment might prove to be overly difficult for some nations due to irregularities in measurement at the lower end of the continuum. There are suggestions that the OECD’s



Longer-Term Strategy of PISA (OECD, 2015) with adaptive testing might provide a better assessment model (Rutkowski, Rutkowski and Liaw, 2019).

In the current context where governments are induced to engage in constant reform in pursuit of global educational targets, a more cautious approach to the understanding of the OECD and its post-2015

education agenda is suggested (Li and Auld, 2020). Interpretation of PISA data is carried out by three primary actors: the OECD, governments and media. Grey and Morris (2018, p. 109) note the role the media plays in creating and publicizing narratives derived from the PISA results, calling it 'mediatized global governance'. PISA is currently considered to be a good proxy measure for the educational quality of a nation, often driving the education policy

PISA data and education policy and reforms within nations are simultaneously accompanied by the emergence of a connected network of economy, commerce and educators.

decision-making with an aim of maximizing economic success. The growing advocacy between PISA data and education policy and reforms within nations is simultaneously accompanied by the emergence of a connected network of economy, commerce and educators, who use PISA data as evidence for ‘best practices’ and gather support to finance their education based on commercial ventures (Auld and Morris, 2014, 2016). The emergence of the OECD as a form of neoliberal educational governance used by policy-makers across nations to selectively drive their agendas within educational reforms is well documented and critiqued (Woodward, 2004; Mahon and McBride, 2009; Ball, 2012; Sellar and Lingard, 2013, 2014; Sjøberg, 2016; You and Morris, 2016; Yasukawa, Hamilton and Evans, 2017).

The authority of an international organization such as the OECD has both ‘rational-legal and moral dimensions’, which is consistent with the observation that an ‘international organization’s political authority is at its zenith when the rational/technical agenda aligns with prevailing

social values and sentiments’ (Eccleston, 2011, p. 248). In this aspect, the enhanced significance of PISA can be explained in terms of new demands for international comparative measures of educational performance in an age of accountability and audit culture (Hopmann, 2008) and evidence-informed policy-making (Head, 2008; Wiseman, 2010). The enhanced significance of data in the OECD’s education policy work has affected modes of educational governance in many societies. Jakobi and Martens (2010) argue that the demand for technical expertise has enabled the organization to enlarge its ‘toolbox of governance mechanisms’. The OECD now produces what we see as globalized education policy discourses (Lingard and Rawolle, 2011), through peer reviews of policy, data generation and analysis, and its impact on the framing and stages of policy-making and enactment within nations. The technical expertise of the OECD has contributed to the emergence of a global education policy field and its intergovernmental structure helps to establish this network of policy

actors across national capitals (Lingard and Rawolle, 2011).

In 2013, the Central Committee of the Communist Party of China announced a series of educational reforms designed to change the disadvantages of a system where

‘one’s fate is determined by one examination’ (Decision, 2013, p. 12) – an often quoted saying regarding the dominant role of the NCEE in China. According to policy discourse, reform of test-centric education is essential for the development of student’s

The technical expertise of the OECD has contributed to the emergence of a global education policy field and its intergovernmental structure helps to establish this network of policy actors across national capitals.

## BOX 2: CHINA

individual talent, innovation and creativity, skills necessary for China’s modernization and knowledge economy. Although varied across provinces, reforms are national; pilot programmes in Shanghai and Zhejiang, in 2014 were later implemented in Beijing, Tianjin, Shandong and Hainan in 2017. In 2018, the reform expanded to eight provincial-level regions, namely Hebei, Liaoning, Jiangsu, Fujian, Hubei, Hunan, Guangdong and Chongqing. In particular, two reforms have been hailed as milestones in improving both educational quality and furthering ‘quality’ (sushi) education goals (Zhen, 2017): student choice of

subjects across previous set streams of fine arts or science, and alternative assessments known as comprehensive quality evaluations (*zonghesuzhipingjia*).

According to policy, the ending of subject streams and the promotion of student choice across six subjects of physics, chemistry, biology, geography, history and politics will cultivate diversified talent by creating more ‘personalized’ education reflective of individual interests and strengths. However, research on student choice has shown an overwhelming preference for science courses among both



students and parents; entrance into Tier 1 universities requires exams in physics or chemistry, and graduating from top universities translates into better job opportunities. There are pedagogical considerations as well, as science majors express a preference for learning outside of rote memorization, a form of pedagogy most associated with fine arts (Frame, 2020).

Combined with seventy years of educational policies and higher education funding geared toward science and technology in the service of national development, the promotion of ‘personalized’ learning runs counter to both historical and institutional arrangements, particularly in a decentralized system where funding for schools is essentially tied to test scores and university admittance. Finally, maximization of one’s NCEE scores has also been shown to influence student choice much more than personal interest (Li, 2017; Tan and Ng, 2018; Frame, 2020). At present, the cultivation of student choice within a system monopolized

by success in the NCEE, school ranking and higher education opportunity remains a myth.

The importance of scores also figures large in comprehensive quality evaluations, a new university admissions requirement based on alternative and formative assessment rather than the conventional and summative assessment of the NCEE (Tan and Ng, 2018). Designed to promote development of students’ moral, intellectual and aesthetic qualities, evaluations include assessments in ideology, moral character, physical and mental health, artistic accomplishment and social service (Deepening, Section B:2). Considering the objective, quantifiable nature of the NCEE, where an anonymous score decides future opportunity, the subjective nature of teacher assessment is a serious roadblock. Issues of fairness and the prevalent use of *guanxi* in Chinese society has led to fears of corruption in the university admission procedure, particularly amongst disadvantaged rural residents

(Liu et al., 2012). In this light, fairness has little to do with issues of equity; rather, the ‘fairness’ of the NCEE lies in its ability to control higher education opportunity through transparency. Meanwhile, the promotion of meritocracy in educational discourse since China’s Opening Up (Vickers and Zeng, 2017), has thoroughly embedded the idea of hard work and rote memorization in education, especially among rural populations (Wang and Ross, 2010). However, resistance is also prevalent amongst urban middle-class children, who face increased academic pressure, spurred by a taken-for-granted belief that increased credentials equals an increase in career opportunity (Liu, 2008). This is compounded by ‘sea turtles’ (*haigui*), a euphemism for overseas graduates who return to China and are highly sought after by top companies; stories of exorbitant spending on summer programmes designed to give children an advantage over ‘sea turtles’ abound on websites such as

Sina Gaokao. Combined with growing unemployment for college graduates, a desire by urban parents to ‘guard’ their top position against increasing rural migration, and the NCEE being the only means for the middle class to compete with elite education, it is doubtful that alternative assessments will be embraced by a majority of urban schools either (Mok and Jiang, 2017).

In conclusion, the capacity to reform the pattern whereby ‘one’s fate is determined by one examination’ is interwoven with widely-held structures and beliefs about the purpose of education, social mobility, even epistemology. Policy-makers and researchers would benefit greatly from recognizing the impact of both sociocultural and broader socio-economic changes on reform implementation in China’s NCEE system.

## BOX 3: AFRICA

Within the African context, there has been an increase in the number of countries engaging with ILSAs, as evidenced by the participation of Senegal and Zambia in PISA-D. The increased involvement of some African countries in ILSAs is driven by the shift in global focus from educational provision to the improvement and measurement of educational quality (**Braslavsky, 2005**). Besides, there is a growing emphasis on the concept of the development of human capital, as measured by learning assessments, being related to a country's economic growth (**Hanushek and Kimko, 2000**).

At the level of national assessment, a rapid growth is also discernible. Statistics show that, since the Dakar conference in 2000, almost 40 per cent of sub-Saharan Africa countries have conducted at least one national assessment, compared to about 25 per cent before 2000. However, together with central Asia, the region still exhibits the lowest level of system-level assessment (**Dakar Framework for Action, 2000**). It is currently almost impossible to find comprehensive, reliable data on the costs of introducing and running a national assessment in most sub-Saharan African countries. It seems that all too often, no proper budgetary planning is done, and accounting records are incomplete.

National assessments (via the information that is generated) have the potential to identify practices that may be responsible for underperformance. Also critical is how information

obtained is utilized to impact education reform in general, and improve learning outcomes, in particular (**Schiefelbein and Schiefelbein, 2003**). For example, the underuse of the available

data is one of the shortcomings of national assessment in many African countries. In a study on an appropriate assessment models for higher education, specifically health sciences and technology, Friedrich-Nel, De Jager and Nel (2005, pp. 881–883) investigated current educational practices characteristic of higher education, concluding that for most of the twentieth century, teaching in higher education was geared to exposing students to masses of facts up to the point where the facts became unmanageable. They concurred with Olivier (1999, p. 69) that written examinations, traditionally associated with content-based education and training, remain the dominant form of assessment used in higher learning institutions in South Africa.

A future perspective suggests it is imperative for African policy-makers to put in place mechanisms for developing multicultural assessment standards (Sedlacek, 1994), and/or redefine learning and assessment practices within the Indigenous worldviews

to equip Indigenous students with relevant employability skills. Hence, indigenization of education and assessment in Africa may effectively facilitate teaching and learning in schools, thereby making education more meaningful and responsive to the needs of Africans in this era of globalized education (Obanya, 2007).

There is no doubt that in terms of resource endowments, both China and Africa possess an abundance of human and material resources. However, in the assessment area, it may be apposite to stress that the indigenous languages and cultural values that are reflective in the educational systems of both could be entrenched in assessment and/or taken into consideration. One of the similarities that underlies educational systems' assessment mechanisms in both China and Africa is not unconnected with advocacy for cultural values in their educational assessments. Regardless, the contrast is reflected in the fact that China is represented as a single strong state in which the Chinese indigenous language is rooted mostly in



Mandarin, whereas there is no such common state or unified language in Africa. Although some scholars have argued for Swahili's adoption as an official language in Africa (Ngugi, 1986; Amidu, 1995; Karenga, 1997; Tabb, 2006), the project is yet to be implemented as Africa is multi-ethnic. Therefore, it is essential to note that one of the purposes of language unification is to facilitate communication among the people in the continent and aid the patterns of assessment.

9.5 .5

## GLOBAL ASSESSMENT AND SUSTAINABILITY

At the 70th Session of the UN General Assembly in September 2015, member states adopted the 2030 Agenda for Sustainable Development (UN, 2015). It aims to engage the nations of the world in

... OECD has sought to appropriate and reinterpret the sustainability agenda by developing metrics for monitoring performance in the domain of ESD and generating related rankings, data and indicators.

collectively promoting sustainable development, decreasing global inequalities and realizing universal quality education. At the heart of the agenda are seventeen SDGs, including SDG 4, which covers education seeking to 'ensure inclusive and equitable quality education and promote lifelong learning opportunities for all'. SDG 4.7 highlights that by 2030 all learners should have the knowledge and skills needed to promote sustainable development, including through ESD and sustainable lifestyles, human rights, gender equality, promotion of a culture of peace and non-violence, global citizenship, and appreciation of cultural diversity and of culture's contribution to sustainable development.

It should be acknowledged that multilateral organizations such as the OECD, the World Bank

and UNESCO have struggled to assert their authority in interpreting and measuring SDG 4 (especially SDG 4.7). As a case in point, the OECD has made efforts to lay claim to special expertise in measuring ESD through its discourse of 'global competencies', repackaging its human capital narrative while effectively marginalizing UNESCO's humanist perspective. With its 'Future of Education and Skills 2030' programme, and the '2030 Learning Compass', the OECD has sought to appropriate and reinterpret the sustainability agenda by developing metrics for monitoring performance in the domain of ESD and generating related rankings, data and indicators (OECD, 2016) (see also W62-ch1).



# 9.6

## Conclusion and discussion

This chapter shows that key arguments about the purpose and nature of educational assessment are not new. Tensions between the formative and summative functions of assessment are as old as formal education itself. Formative assessment involves the pedagogical skill of monitoring students' learning in order to

identify learning needs and adjust teaching appropriately. Such ongoing assessment for learning is valued for enhancing teachers' focus on the needs of their own students and for achieving greater equity of student outcomes. Research has indicated that an emphasis on formative assessment tends to be associated with more



... an emphasis on formative assessment tends to be associated with more clarity in the setting of learning objectives, more variation in instructional practices and a higher level of student interactions in the classroom.

clarity in the setting of learning objectives, more variation in instructional practices and a higher level of student interactions in the classroom (Bennett, 2011). Regarding assessment itself as a form of learning encourages students to become independent and confident learners. The role of teachers then is to help students cultivate self-assessment skills and a growth mindset by creating environments where they are encouraged to confront challenges, while ensuring resources such as models of good practice and emotional support are readily available (for teachers as well as students). But all this assumes high levels of teacher autonomy and professionalism, which in many societies are lacking, or even consciously obstructed by systems for controlling and monitoring teachers (WG2-ch10). In contexts where control over education and teachers is prioritized – whether due to an autocratic political environment, or in the name of neoliberal ‘accountability’ – summative forms of assessment, measuring student achievement according to externally imposed

‘outcomes’ metrics, tend to predominate.

Both summative and formative approaches have their place, since ultimately students’ learning will require formal certification to enable them to move on to the next stage of education or into the workforce. However, there is a clear tension between assessment or evaluation as an inescapable and necessary feature of any learning process or education system, and the dangers of excessive emphasis on outcomes-focused assessment. Echoing Dore’s (1976) critique of the excessive reliance on credentials in many modern education systems, there is a growing literature today on the phenomenon of ‘meritocracy’ and its relationship with social, political and cultural contexts (see WG2-ch1, WG2-ch3). When educational assessment is analysed in context, we are not just discussing better or worse ways of measuring learning from a technical standpoint, but also who is measuring, why and how the results of those measurements are used. The increasingly intense



An important question then is how current research on assessment may contribute to the development of forms of assessment conducive to human flourishing within sustainable and equitable societies.

focus on educational ‘outcomes’ in many societies over recent years, driven in part by the influence of the OECD, is related to assumptions about how society and the economy should work – in particular, the idea that ‘merit’ defined in terms of measurable educational outcomes should determine life chances, as well as assumptions that education’s purpose primarily involves human capital generation instead of human flourishing and the promotion of truly sustainable and equitable societies.

An important question then is how current research on assessment may contribute to the development of forms of assessment conducive to human flourishing within sustainable and equitable societies. There is a risk of reducing the notion of flourishing to what can be technically measured in large-scale assessment exercises, resulting in narrow and distorted conceptions. A focus on flourishing implies a holistic and humanistic view of education built around a model of assessment for and as

learning. It also implies that some important aspects of education for child flourishing cannot be appropriately assessed given our current technical capabilities and would be better left out of assessments. And in that case, the political and normative issue is not assessing child flourishing, but rather what assessment ‘mix’ is most compatible with an approach to education that balances its instrumental functions with its intrinsic importance to a fulfilling life. Some new developments of educational assessment may be considered helpful, including assessments of higher-order skills (such as problem-solving and collaboration), application of advanced technology and improving teaching through assessment.

Understanding the complex relations between assessment and context requires analysis of public discourse on science and evidence, the influences that shape it and the role of vested interests (see WG2-ch1). Amongst the choices relating to assessment are choices regarding what counts

as ‘evidence’. Here we need to address a number of questions concerning issues of science, evidence and objectivity, which are at the heart of much public debate over assessment and education more generally around the world today. Given the complexity of such issues, the extent to which it is possible or desirable to aspire to normative or prescriptive ‘solutions’ to assessment-related problems is open to question. When we consider the role that assessment can or should play in improving our education systems, we need to remind ourselves of the limits of the capacity of assessment reform to achieve the desired improvements. Calls for the introduction of ever more sophisticated and intrusive forms of summative assessment, or of techniques combining formative and summative functions, are often attributable to the widespread tendency in many societies today to see education as the ‘silver bullet’ for a variety of social problems. The heightened stakes thereby associated with the work of teachers and schools generates intensified pressure

for ‘accountability’, which can ultimately subject teachers to ever more intrusive forms of command and control, diminishing their professional status and cramping their autonomy (see **WG2-ch10**).

Assessment in crucial ways both influences, and is influenced by, the political and ideological context. A particularly prominent feature of that context in the early twenty-first century is the conduct of standardized global assessments (e.g. PISA), which have acquired such high stakes for many governments, in turn helping to spur a proliferation of national testing regimes. Apart from its implications for teachers’ autonomy and status, this risks narrowing the curricular focus, as educational officials, teachers, parents and students themselves reorient learning to the maximization of test scores. In such circumstances, improvements in scores can be seen not necessarily as the result of improvements in learning, but rather of improvements in test preparation – in teachers’ and

... the extent to which it is possible or desirable to aspire to normative or prescriptive ‘solutions’ to assessment-related problems is open to question.

students' skill in 'gaming' the test. Moreover, as noted in **WG2-ch3**, a vast and rapidly expanding global industry of supplementary private tutoring has emerged in recent decades, overwhelmingly geared to coaching students for high-stakes public examinations.

While this chapter has primarily focused on assessment in context, the expansion and intensification of testing in many societies means that metrics have often become an end in themselves. The use of quantitative metrics to judge, rank and monitor performance across a whole range of public institutions, but especially in schools and colleges, has become embedded in the educational systems of many societies, especially those most profoundly influenced by the tenets of NPM. The resulting 'tyranny of metrics' (**Muller, 2018**), justified in the language of public 'accountability', can be hard to resist: why should anyone object to more transparency? What have they got to hide? But time that teachers or other professionals spend filling in forms or administering externally mandated

tests detracts from time available for the exercise of autonomous professional judgement (applied amongst other things, to devising forms of assessment that teachers themselves may consider useful or appropriate); similarly, participation in ILSAs (e.g. PISA-D) also risks diverting scarce resources available in poor nations from other more important priorities. Indeed, the emphasis on ever more stringent accountability and transparency clearly implies an absence of trust in teachers' skills and professionalism that many have found profoundly demoralizing. This lack of trust in teachers and denial of their agency are starkly at odds with the proclaimed commitment of transnational institutions or national ministries to fostering confidence, autonomy and dignity in learners. We may wonder how teachers can be expected to model or inculcate qualities deemed essential to flourishing, when these are increasingly denied in their own professional lives. These issues are further discussed in **WG2-ch10**.

We may wonder how teachers can be expected to model or inculcate qualities deemed essential to 'flourishing, when these are increasingly denied in their own professional lives.

... instrumentalist discourses of human capital and accountability versus humanistic conceptions of the intrinsic value of education in promoting flourishing.

What, then, are the key assumptions informing global approaches to assessment today? Assumptions concerning the nature of education or learning, its purpose, and the role of assessment need to be explicated. There is the need to define the ideological beliefs, economic interests and political agendas that are driving reforms to assessment regimes around the world. Key considerations here include instrumentalist discourses of human capital and accountability versus humanistic conceptions of the intrinsic value of education in promoting flourishing. Despite the joint articulation of priorities in the 2030 Agenda set by the UN, it is important to note that UNESCO and the OECD propose very different normative frameworks for reaching these goals, especially regarding expected educational outcomes, and how they are to be assessed (see **Vaccari and Gardinier, 2019**). UNESCO embodies a legacy of humanistic and emancipatory ideals, as witnessed in documents such as the **Delors Report** (International Commission on Education for the

Twenty-first Century, 1996), and the more recent report, **Rethinking education: towards a global common good education: towards a global common good?** (UNESCO, 2015). Their focus is on global justice and equality, with a strong emphasis on values and rights, assessed through periodic monitoring reports that draw on a range of quantitative and qualitative data (see UNESCO, 2020 for latest report). The OECD, on the other hand, follows a more technocratic and economic approach, based on an underlying theory of human capital. In terms of outcomes, the OECD places strong emphasis on measurable indicators framed in terms of 'skills'.

While both UNESCO and the OECD promote the global priorities put forward in the 2030 Agenda, an underlying question is whether their respective worldviews are compatible in ensuring educational outcomes that promote more sustainable futures. Elfert (2017) argues that UNESCO's rights based approach has gradually been displaced by the 'hegemony of the economic



worldview' of the OECD. And there is no doubt that the technical and economic prowess of the OECD has seen it play an increasingly prominent role in promoting metrics for assessing ESD. However, considering the loosely defined nature of sustainability, and increasing political polarization around the world, the humanistic and emancipatory vision of UNESCO is important in bringing a critical perspective to the debate over the role of educational assessment in promoting greater social and environmental justice in a global context where the neoliberal outlook remains widely entrenched.

9.6 .1

## KEY MESSAGES

*Assessment is a vital component of the learning process*, but it should be applied with great caution. Excessive focus on measurement of pupil achievement can lead to a narrowing of curricular focus, intolerable pressure on teachers and learners, and distortion of the learning process. A greater usage of the forms of assessment for and as learning may help address this imbalance.

*Whilst technological change facilitates our understanding of assessment*, including the potential for expanding and refining assessment techniques to make it more adaptive to individual learning requirements, some important features of education for human flourishing cannot easily, or appropriately, be subjected to quantitative measures.

*Moreover, the resulting enhancement of our capacity to test* and measure learning in new ways also carries significant risks – political and social, as well as educational. Tools for the sophisticated assessment of individual learning can often also be applied for purposes of surveillance and control.

*The global intensification of assessment is signalled by the continuing expansion of large-scale assessments*, now extending into more low- and middle-income nations. International organizations, such as the OECD, UNESCO and the World Bank, are the key promoters and

suppliers often with the help of technical partners.

*PISA, perhaps the most well-known example of an ILSA*, is widely perceived as offering insights into the relationship between a country's educational outcomes and its economic growth prospects. However, this relationship is far more complex and uncertain than is commonly recognized, and PISA's popularity illustrates the dangers of excessive quantitative measurement of learning and an overly instrumentalist vision of 'education-as-human capital generation'. Related risks include a narrowing curricular focus, increasing competitive intensity and the persistence of unsustainable economic behaviour.

*We therefore need to pay close attention to who conducts assessment and for what purposes*. Assessment should serve a vision of teaching and learning that respects the agency and dignity of teachers and learners; it should not become a tool of

state oppression or an instrument primarily for maximizing efficiency in the generation of ‘human capital’. In assessing many of the most important areas of learning, digital technology is no substitute for the professional judgement of experienced educators.

### SYNTHESIS OF THE PERSPECTIVES FROM NEUROSCIENCES AND SOCIAL SCIENCES

*Evidence from work in the learning sciences and neurosciences* shows that formative assessment and assessment as learning throughout the course of schooling can have positive effects on academic achievement.

*However, high-stakes, summative assessment regimes can distort the learning process* and cause social and psychological harm. The risks of excessively intense measurement of student achievement are illustrated by the experience of societies (for example across much of Asia and, increasingly, in the

anglophone West) where high-stakes assessment regimes have become especially embedded. These risks include not only extremely intense educational competition (credentialism), but also the reinforcing of social divisions based on the assumption that ‘meritocracy’ may justify inequality.

*ILSAs, such as PISA, have contributed to harmonizing the global assessment landscape and the formation of a global education policy field.* The enhanced significance of PISA derives legitimacy from what is portrayed as a scientific approach to data gathering. It can also be partially explained in terms of new demands for international comparative measures of educational performance in an age of accountability and audit culture. However, the role of ideology and vested interests in promoting the kind of accountability and audit culture associated with dominant assessment regimes is insufficiently appreciated or understood by policy-makers, media and the general public.

*Many extremely valuable forms of learning are hard to measure quantitatively,* and the expectation that any worthwhile learning must be subjected to quantitative measurement can lead to neglect of some of the most important curricular areas.

*Quantitative assessment is especially difficult in curricular areas associated with cultural, historical, artistic, political, ethical or 'values' education* – including social and emotional learning (SEL). Some research in EN has supported calls for SEL to be integrated within the larger assessment framework if the educational world is to move towards the goals of sustainable development and human flourishing. But quantitative metrics for SEL remain elusive and attempts to design them problematic. (see **WG2-ch8 for a discussion** of other problems with SEL discourse).

*Much work in EN has focused on tracking and analysing 'individual differences' in learning.* Neuroscience-informed formative assessment tools can help to identify differences in students' learning processes and thus potentially aid efforts to support individual learning needs. However, it does not follow that neuroscientifically-informed assessment alone can promise equitable, individually tailored learning opportunities for all students.

*As with other aspects of education, the COVID-19 pandemic and the related shift to online learning has provided insights into modes of assessment,* particularly both the potential and the limitations of information technology in supporting and assessing learning. Attempts to substitute algorithms for conventional public examinations have reminded us that algorithms are poor tools for predicting future performance, and only as reliable as the information on which they are based – information that is ultimately selected by fallible human actors.



**9.6 .2****KEY  
RECOMMENDATIONS**

For researchers and practitioners,  
understanding the complex

relationship between assessment  
and context requires an  
understanding of how ideological  
beliefs, economic interests and  
political agendas drive assessment  
reforms around the world.

Governments and other stakeholders  
should be aware that extending the  
scope of educational assessment for

its own sake is not necessarily a good thing, and that excessively intense or intrusive assessment regimes can have seriously harmful effects – on individual learners, on teachers and on society more broadly.

Governments, in consultation with other stakeholders, are encouraged to specify at the outset the purpose of any proposed assessment reform and the problems or issues it is intended to address.

It is especially important that teachers be centrally involved in decisions over reforming assessment practices. Failure to do so will exacerbate problems of teacher motivation and deprofessionalization.

Educators and policy-makers should, as far as possible, ensure that policy and practice are informed by the findings of relevant research – but in seeking expert advice, it is crucial to consult social scientists (who study education systems in their social, political and cultural context) as well as natural scientists (who may offer insights into the neural or biological

mechanisms related to the learning process).

Decisions over assessment reform need to take careful account of the diversity of our societies and cultures, and beware of the ways in which assessment regimes can unjustly privilege particular cultural traditions while marginalizing or suppressing others.

Educational assessment can be improved via development of instruments and statistical methods in the future. Yet in deciding whether, or how far, to deploy new assessment tools informed by neuroscientific research or supported by new technology, educators and policy-makers should proceed with caution, taking care to balance considerations such as the privacy, autonomy and agency of learners and teachers, the (often unintended) impact that new forms of assessment can have on the focus and content of learning, and cost.

# REFERENCES

- Adams, R. and Cresswell, J. (2016). PISA for development technical strand A: enhancement of PISA cognitive instruments. Paris: OECD.
- Adams, R.J. and Gonzalez, E.J. (1996). 'The TIMSS test design', in Martin, M.O. and Kelly, D.L. (eds.) Third International Mathematics and Science Study (TIMSS) technical report, volume I: design and development. Chestnut Hill, MA: Boston College.
- Addey, C. (2015) 'Participating in international literacy assessments in Lao PDR and Mongolia: a global ritual of belonging', in Hamilton, M. and Addey, C. (eds.) Literacy as numbers: researching the politics and practices of international literacy assessment regimes. Cambridge: Cambridge University Press, pp. 147–164.
- Addey, C. and Sellar, S. (2018) 'Why do countries participate in PISA? Understanding the role of international large-scale assessments in global education policy', in Verger, A., Novelli, M. and Altinyelken, H.K. (eds.) Global education policy and international development: new agendas, issues and policies. London: Bloomsbury, pp. 97–117.
- Aggarwal, N.K. (2013) 'Cultural psychiatry, medical anthropology, and the DSM-5 field trials', *Medical Anthropology*, 32(5), pp. 393–398.
- Alexander, R.J. (2000) *Culture and pedagogy: international comparisons in primary education*. Oxford: Blackwell.
- Amidu, A.A. (1995) 'Kiswahili, a continental language: how possible is it? (Part 1)', *Nordic Journal of African Studies*, 4(2), pp. 50–70.
- Andrade, H. and Du, Y. (2007). 'Student responses to criteria-referenced self-assessment', *Assessment & Evaluation in Higher Education*, 32(2), pp. 159–181.
- Apple, M.W. (1993). *Official knowledge: democratic education in a conservative age*. New York: Routledge.
- Asen, R., Gurke, D., Conners, P., Solomon, R. and Gumm, E. (2013). 'Research evidence and school board deliberations: lessons from three Wisconsin school districts', *Educational Policy*, 27(1), pp. 33–63.
- Auld, E. and Morris, P. (2014). 'Comparative education, the "new paradigm" and policy borrowing: constructing knowledge for educational reform', *Comparative Education*, 50(2), pp. 129–155.
- Auld, E. and Morris, P. (2016). 'PISA, policy and persuasion: translating complex conditions into education "best practice"', *Comparative Education*, 52(2), pp. 202–229.
- Auld, E., Rapple, J. and Morris, P. (2019) 'PISA for development: how the OECD and World Bank shaped education governance post-2015', *Comparative Education*, 55(2), pp. 197–219.
- Ball, S.J. (2012) *Global education inc: new policy networks and the neo-liberal imaginary*. New York: Routledge.
- Barch, D.M. and Yarkoni, T. (2013). 'Introduction to the special issue on reliability and replication in cognitive and affective neuroscience research', *Cognitive, Affective, & Behavioral Neuroscience*, 13(4), pp. 687–689.
- Becker, G.S. (1964). *Human capital: a theoretical and empirical analysis, with special reference to education*. New York: National Bureau of Economic Research.
- Beech, J. and Rizvi, F. (2017) 'Revisiting Jullien in an era of globalisation', *Compare: A Journal of Comparative and International Education*, 47(3), pp. 374–387.
- Benavot, A. and Tanner, E. (2008) 'The growth of national learning assessments in the world, 1995–2006', Background paper for the Education for All global monitoring report. Paris: UNESCO.
- Bennett, R.E. (2011). 'Formative assessment: a critical review', *Assessment in Education: Principles, Policy & Practice*, 18(1), pp. 5–25. doi:10.1080/0969594X.2010.513678.
- Bennett, R.E. (2015). 'The changing nature of educational assessment', *Review of Research in Education*, 39(1), pp. 370–407.



- Bennett, R.E. and Gitomer, D.H. (2009). 'Transforming K–12 assessment: integrating accountability testing, formative assessment and professional support', in Wyatt-Smith, C. and Cumming, J.J. (eds.) *Educational assessment in the 21st century: connecting theory and practice*. Amsterdam: Springer, pp. 43–61.
- Black, M.M., Lutter, C.K. and Trude, A.C. (2020) 'All children surviving and thriving: re-envisioning UNICEF's conceptual framework of malnutrition', *The Lancet Global Health*, 8(6), pp. e766–e767.
- Black, P. and Wiliam, D. (1998). 'Assessment and classroom learning', *Assessment in Education: Principles, Policy & Practice*, 5(1), pp. 7–74.
- Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2004). 'Working inside the black box: assessment for learning in the classroom', *Phi Delta Kappan*, 86(1), pp. 8–21.
- Boud, D. and Falchikov, N. (2007). 'Developing assessment for informing judgement', in Boud, D. and Falchikov, N. (eds.) *Rethinking assessment in higher education: learning for the longer term*. London: Taylor & Francis, pp. 181–197.
- Bowers, J.S. (2016). 'The practical and principled problems with educational neuroscience', *Psychological Review*, 123(5), pp. 600–612.
- Braslavsky, C. (2005). 'The history of education and the contemporary challenge of quality education for all', *Prospects*, 35(4), pp. 385–399.
- Britto, P.R., Cerezo, A. and Ogbunugafor, C.B. (2008) 'How evidence can be used to inform policy: a case study of early childhood evidence-based policy development', *International Journal of Early Childhood*, 40(2), pp. 101–118.
- Broadfoot, P., Daugherty, R., Gardner, J., Harlen, W., James, M. and Stobart, G. (2002) *Assessment for learning: 10 principles*. Cambridge: University of Cambridge School of Education.
- Brock-Utne, B. (2016) 'The Ubuntu paradigm in curriculum work, language of instruction and assessment', *International Review of Education*, 62, pp. 29–44.
- Bruer, J.T. (1997) 'Education and the brain: a bridge too far', *Educational Researcher*, 26(8), pp. 4–16.
- Bruer, J.T. (1999) *The myth of the first three years: a new understanding of early brain development and lifelong learning*. New York: Simon and Schuster.
- Bruer, J.T. (2016) 'Where is educational neuroscience?', *Educational Neuroscience*, 1. doi: <https://doi.org/10.1177/2377616115618036>.
- Bryk, A.S. and Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. New York: Sage Publications.
- Bunge, S. and Souza, M. (2009) 'Executive function and higher-order cognition: neuroimaging', *Encyclopedia of Neuroscience*, 4, pp. 111–116.
- Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., ... and Munafò, M.R. (2013) 'Power failure: why small sample size undermines the reliability of neuroscience', *Nature Reviews Neuroscience*, 14(5), pp. 365–376.
- Caine, R.N. and Caine, G. (1991) *Making connections: teaching and the human brain*. New York: Addison-Wesley.
- Carney, S., Rappleye, J. and Silova, I. (2012). 'Between faith and science: world culture theory and comparative education', *Comparative Education Review*, 56(3), pp. 366–393.
- Carr-Hill, R. (2015) *PISA for development technical strand C: incorporating out-of-school 15-year-olds in the assessment*, OECD Education Working Papers, No. 120. Paris: OECD.
- Central Committee of the CCP. (2013). *Decision of the CCCPC on some major issues concerning comprehensively deepening the reform and comprehensively promoting quality education*. Available at: [http://www.china.org.cn/chinese/2014-01/17/content\\_31226494.htm](http://www.china.org.cn/chinese/2014-01/17/content_31226494.htm) (Accessed: 04 February 2021).

# REFERENCES

- Chiao, J.Y. (2015) 'Current emotion research in cultural neuroscience', *Emotion Review*, 7(3), pp. 280–293.
- Chiao, J.Y., Iidaka, T., Gordon, H.L., Nogawa, J., Bar, M., ... and Ambady, N. (2008) 'Cultural specificity in amygdala response to fear faces', *Journal of Cognitive Neuroscience*, 20(12), pp. 2167–2174.
- Clark, H., Coll-Seck, A.M., Banerjee, A., Peterson, S., Dalglish, S.L., ... and Claeson, M. (2020). 'A future for the world's children? A WHO–UNICEF–Lancet Commission', *The Lancet*, 395(10224), pp. 605–658.
- Committee for Children (2019). What is social-emotional learning? Available at: <https://www.cfchildren.org/policy-collateral/one-pagers/what-is-social-emotional-learning/> (Accessed: 30 January 2021).
- Dahl, C.J., Wilson-Mendenhall, C.D. and Davidson, R.J. (2020) 'The plasticity of well-being: a training-based framework for the cultivation of human flourishing', *Proceedings of the National Academy of Sciences of the United States of America*, 117(51), pp. 32197–32206.
- Dakar Framework for Action (2000). Education for all: meeting our collective commitments, World Education Forum Dakar, Senegal, 26–28 April. Available at: [https://sdgs.un.org/keyword=content&f%5B0%5D=content\\_type%3Adocuments](https://sdgs.un.org/keyword=content&f%5B0%5D=content_type%3Adocuments) (Accessed: 30 January 2021).
- De Ayala, R.J. (2013) *The theory and practice of item response theory*. New York: Guilford Press.
- Dewey, J. (1923) *Democracy and education: an introduction to the philosophy of education*. New York: Macmillan.
- Dirkx, J.M. (2008). 'The meaning and role of emotions in adult learning', *New Directions for Adult and Continuing Education*, 2008(120), pp. 7–18.
- Dore, R. (1976). *The diploma disease: education, qualification and development*. London: Allen and Unwin.
- Dresser, R. (2012). 'Reviving oral reading practices with English learners by integrating social-emotional learning', *Multicultural Education*, 20(1), pp. 45–50.
- Dunlosky, J. and Nelson, T.O. (1992) 'Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect', *Memory & Cognition*, 20(4), pp. 374–380.
- Dweck, C.S. (2008) *Mindset: the new psychology of success*. New York: Ballantine Books.
- Eccleston, R. (2011) 'The OECD and global economic governance', *Australian Journal of International Affairs*, 65(2), pp. 243–255.
- Edwards, S. (2019) 'India's re-entry to PISA triggers mixed response', *Devex*, 20 February. Available at: <https://www.devex.com/news/sponsored/india-s-re-entry-to-pisa-triggers-mixed-response-94286> (Accessed: 3 March 2021).
- Elfert, M. (2017) *UNESCO's utopia of lifelong learning: an intellectual history*. New York: Routledge.
- Engel, L.C., Rutkowski, D. and Thompson, G. (2019) 'Toward an international measure of global competence? A critical look at the PISA 2018 framework'. *Globalisation, Societies and Education*, 17(2), pp. 117–131.
- Falchikov, N. (2013). *Improving assessment through student involvement: practical solutions for aiding learning in higher and further education*. London: Routledge.
- Farley-Ripple, E., May, H., Karpyn, A., Tilley, K. and McDonough, K. (2018). 'Rethinking connections between research and practice in education: a conceptual framework', *Educational Researcher*, 47(4), pp. 235–245.
- Farstrup, A.E. and Samuels, S.J. (2002) *What research has to say about reading instruction*. Newark, DE: International Reading Association.
- Feiler, J.B. and Stabio, M.E. (2018) 'Three pillars of educational neuroscience from three decades of literature', *Trends in Neuroscience and Education*, 13, pp. 17–25.



- Feniger, Y. and Atia, M. (2019) 'Rethinking cause and effect: analysing economic growth and PISA scores over a period of 15 years', in Gorur, R., Sellar, S. and Steiner-Khamsi, G. (eds) *World yearbook of education 2019: comparative methodology in the era of big data and global networks*. London: Routledge, pp. 104–119.
- Ferreira, M., Martinsone, B. and Talić, S. (2020) 'Promoting sustainable social emotional learning at school through relationship-centered learning environment, teaching methods and formative assessment', *Journal of Teacher Education for Sustainability*, 22(1), pp. 21–36.
- Fischer, G.H. and Molenaar, I.W. (eds.) (2012) *Rasch models: foundations, recent developments, and applications*. New York: Springer Science & Business Media.
- Florensa, C., Duan, Y. and Abbeel, P. (2017) 'Stochastic neural networks for hierarchical reinforcement learning', 5th International Conference on Learning Representations, Toulon, 24–26 April.
- Ford, M. (2013) 'Achievement gaps in Australia: what NAPLAN reveals about education inequality in Australia', *Race Ethnicity and Education*, 16(1), pp. 80–102.
- Frame, M. (2020) 'The institutional pillars of China's National College Entrance Exam: a case study of Gaozhong high school and NCEE reforms' Unpublished doctoral dissertation. University of Massachusetts, Amherst.
- Friedrich-Nel, H.S., De Jager, L. and Nel, M.M. (2005). 'An assessment model in outcomes-based education and training (OBET) for health sciences and technology in South Africa', *South African Journal of Higher Education*, 19(5), pp. 880–899.
- Frith, U., Bishop, D., Blakemore, C., Blakemore, S.J., Butterworth, B. and Goswami, U. (2011) *Brain waves module 2: neuroscience: implications for education and lifelong learning*. The Royal Society. Available at: <https://www.interacademies.org/sites/default/files/publication/4294975733.pdf> (Accessed: 14 December 2021).
- Geake, J. (2008) 'Neuromythologies in education', *Educational Research*, 50(2), pp. 123–133.
- Gibbs, G. (2006) 'Why assessment is changing', in Bryan, C. and Clegg, K. (eds.) *Innovative assessment in higher education*. Abingdon: Routledge, pp. 11–22.
- Gipps, C. (2002). 'Sociocultural perspectives on assessment', in Wells, G. and Claxton, G. (Eds.) *Learning for life in the 21st century*. Oxford: Blackwell, pp. 73–83.
- Gleason, T. and Narvaez, D. (2019) 'Beyond resilience to thriving: optimizing child wellbeing', *International Journal of Wellbeing*, 9(4). doi: <https://doi.org/10.5502/ijw.v9i4.987>.
- Gohl, E.M., Gohl, D. and Wolf, M.A. (2009) 'Assessments and technology: a powerful combination for improving teaching and learning', in Pinkus, L.M. (ed.) *Meaningful measurement: the role of assessments in improving high school education in the twenty-first century*. Washington, DC: Alliance for Excellent Education, pp. 183–197.
- Green, A. (1997) *Education, globalization and the nation state*. Basingstoke: Macmillan.
- Grek, S. (2009) 'Governing by numbers: the PISA "effect" in Europe', *Journal of Education Policy*, 24(1), pp. 23–37.
- Grey, S. and Morris, P. (2018). 'PISA: multiple "truths" and mediated global governance', *Comparative Education*, 54(2), pp. 109–131.
- Grigorenko, E.L. and Sternberg, R.J. (1998). 'Dynamic testing', *Psychological Bulletin*, 124(1), pp. 75–111.
- Grotlüschen, A. (2018) 'Global competence – does the new OECD competence domain ignore the global South?', *Studies in the Education of Adults*, 50(2), pp. 185–202.

# REFERENCES

- Gu, M., Ma, J. and Teng, J. (eds.) (2017) 'The Gaokao experience of Chinese students', *Portraits of Chinese schools*. Singapore: Springer, pp. 59–75.
- Han, S. and Northoff, G. (2008) 'Culture-sensitive neural substrates of human cognition: a transcultural neuroimaging approach', *Nature Reviews Neuroscience*, 9(8), pp. 646–654.
- Hanushek, E.A. and Kimko, D.D. (2000) 'Schooling, labour-force quality and the growth of nations', *American Economic Review*, 90(5), pp. 1184–1208.
- Hanushek, E.A. and Woessmann, L. (2008) 'The role of cognitive skills in economic development', *Journal of Economic Literature*, 46(3), pp. 607–668.
- Harris, P., Chinnappan, M., Castleton, G., Carter, J., De Courcy, M. and Barnett, J. (2013) 'Impact and consequence of Australia's national assessment program – literacy and numeracy (NAPLAN): using research evidence to inform improvement', *TESOL in Context*, 23(1/2), pp. 30–52.
- Hayduk, L.A. (1987) *Structural equation modeling with LISREL: essentials and advances*. Baltimore: Johns Hopkins University Press.
- Head, B.W. (2008) 'Three lenses of evidence-based policy', *Australian Journal of Public Administration*, 67(1), pp. 1–11.
- Henrich, J., Heine, S.J. and Norenzayan, A. (2010) 'The weirdest people in the world?', *Behavioral and Brain Sciences*, 33(2–3), pp. 61–83.
- Hinton, C., Miyamoto, K. and Della-Chiesa, B. (2008) 'Brain research, learning and emotions: implications for education research, policy and practice', *European Journal of Education*, 43(1), pp. 87–103.
- Hopmann, S.T. (2008) 'No child, no school, no state left behind: schooling in the age of accountability', *Journal of Curriculum Studies*, 40(4), pp. 417–456.
- Howard-Jones, P.A. (2008) 'Education and neuroscience', *Educational Research*, 50(2), pp. 119–122.
- Huber, D.E., Potter, K.W. and Huszar, L.D. (2019) 'Less “story” and more “reliability” in cognitive neuroscience', *Cortex, a Journal Devoted to the Study of the Nervous System and Behavior*, 113, pp. 347–349.
- Hunsley, J. and Mash, E.J. (2007) 'Evidence-based assessment', *Annual Review of Clinical Psychology*, 3(1), pp. 29–51.
- Ingram, J., Lindorff, A., Sani, N., McCann, E. and Riggall, A. (2020) *TALIS video study: national report*. UK: Department for Education.
- International Commission on Education for the Twenty-first Century (1996) *Learning: the treasure within: Report to UNESCO of the International Commission on Education for the Twenty-first Century*. Paris: UNESCO.
- Jackson, P.W. (1986) *The mimetic and the transformative: Alternative outlooks on teaching*. New York: Teachers College Press.
- Jakobi, A.P. and Martens, K. (2010). 'Expanding and intensifying governance: the OECD in education policy', in Martens, K. and Jakobi, A.P. (eds.) *Mechanisms of OECD governance: international incentives for national policy making*. Oxford Scholarship Online, pp. 163–179.
- Jalongo, M.R. (2008) "“Enriching the brain”: the link between contemporary neuroscience and early childhood traditions", *Early Childhood Education Journal*, 35(6), pp. 487–488.
- Jensen, E. (2005) *Teaching with the brain in mind*. Arlington: ASCD.
- Kaess, K. (2018) 'The cross-Atlantic knowledge divide, or PISA for Development: should one size ever fit all?', *Atlantic Studies*, 15(3), pp. 349–364.
- Kamens, D. (2014) *Globalization, innovation in education policy and practice, and cultural resistance: lessons from international testing*. Paper presented at The Future of Curriculum

- Studies and Research on Schooling, University of Vienna, Vienna, April, pp. 28–30.
- Kamens, D.H. and Benavot, A. (2011) 'National, regional and international learning assessments: trends among developing countries, 1960–2009', *Globalisation, Societies and Education*, 9(2), pp. 285–300.
- Kamens, D.H. and McNeely, C.L. (2010) 'Globalization and the growth of international educational testing and national assessment', *Comparative Education Review*, 54(1), pp. 5–25.
- Karenga, M. (1997). *Kwanzaa: a celebration of family, community and culture*. Los Angeles: University of Sankore Press.
- Karlen, Y. (2016) 'Differences in students' metacognitive strategy knowledge, motivation, and strategy use: a typology of self-regulated learners', *The Journal of Educational Research*, 109(3), pp. 253–265.
- King, R.B. and Chen, J. (2019) 'Emotions in education: Asian insights on the role of emotions in learning and teaching', *The Asia-Pacific Education Researcher*, 28(4), pp. 279–281.
- Klees, S.J. (2016) 'Human capital and rates of return: Brilliant ideas or ideological dead ends?', *Comparative Education Review*, 60(4), pp. 644–672.
- Knowland, V.C.P. (2020). 'Educational neuroscience: ethical perspectives,' in Thomas M.S.C., Mareschal, D. and Dumontheil, I. (eds.) *Educational neuroscience: development across the life span*. London: Routledge, pp. 474–499.
- Koizumi, H. (2004) 'The concept of "developing the brain": a new natural science for learning and education', *Brain and Development*, 26(7), pp. 434–441.
- Komatsu, H. and Rappleye, J. (2017). 'A new global policy regime founded on invalid statistics? Hanushek, Woessmann, PISA, and economic growth', *Comparative Education*, 53(2), pp. 166–191.
- Kort, B., Reilly, R. and Picard, R.W. (2001). 'An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion', in *Proceedings IEEE International Conference on Advanced Learning Technologies*, pp. 43–46.
- Kreft, I.G., De Leeuw, J. and Aiken, L.S. (1995) 'The effect of different forms of centering in hierarchical linear models', *Multivariate Behavioral Research*, 30(1), pp. 1–21.
- Law, N.W.Y. and Tsang, H.W.C. (2019). 'Implications of social neuroscience for learning technology research and development', in Parsons, T.D., Lin, L. and Cockerham, D. (eds.) *Mind, brain and technology: learning in the age of emerging technologies*. New York: Springer International Publishing, pp. 161–176.
- Leksmono, A., Prihandoko, A.C. and Murtikusuma, R.P. (2019). 'Students' creative thinking process in completing mathematical PISA test concerning space and shape', *Journal of Physics: Conference Series*, 1211(1). doi: 10.1088/1742-6596/1211/1/012073.
- Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., ... and Jocelyn, L. (2004) *International outcomes of learning in mathematics literacy and problem solving: PISA 2003 results from the U.S. perspective*. NCES 2005-003. Washington, DC: US Department of Education.
- Lewis, S. (2017) 'Governing schooling through "what works": the OECD's PISA for schools', *Journal of Education Policy*, 32(3), pp. 281–302.
- Li, J. (2017) 'The impact of the reform of the new college entrance examination in Zhejiang Province and its response' *Education and Examination*, 2, pp. 26-31.
- Li, X. and Auld, E. (2020) 'A historical perspective on the OECD's "humanitarian turn": PISA for Development and the Learning Framework 2030', *Comparative Education*, 56(4), pp. 503–521.



# REFERENCES

- Lietz, P., Wagemaker, H., Neuschmidt, O. and Hencke, J. (eds.) (2008). *Educational issues in the Middle East North Africa region: outcomes of the IEA Arab region training seminar series 2006/2007*. Netherlands: The International Association for the Evaluation of Educational Achievement.
- Lingard, B. and Rawolle, S. (2011) 'New scalar politics: implications for education policy', *Comparative Education*, 47(4), pp. 489–502.
- Liu, F. (2008) 'Constructing the autonomous middle-class self in today's China: the case of young-adult only-children university students' *Journal of Youth Studies*, 11(2), pp. 193–212.
- Liu, Y., Wang, Z. and Yang, X. (2012). 'Selecting the elite: status, geography and capital: admission of rural students into Peking University', *International Journal of Chinese Education*, 1, pp. 19–53.
- Lockheed, M. (2013) 'Causes and consequences of international large-scale assessments in developing countries', in Meyer, H.D. and Benavot, A. (eds.) *PISA, power, and policy: the emergence of global educational governance*. Providence, RI: Symposium Books, pp. 185–206.
- Lockheed, M., Prokic-Bruer, T. and Shadrova, A. (2015) *The experience of middle-income countries participation in PISA 2000–2015*. Paris: OECD.
- Mahon, R. and McBride, S. (2009). 'Standardising and disseminating knowledge: the role of the OECD in global governance', *European Political Science Review*, 1(1), pp. 83–101.
- Martens, K. (2007) 'How to become an influential actor – the 'comparative turn' in OECD education policy', in: Martens, K., Rusconi, A. and Leuze, K. (eds.) *New arenas of education governance: the impact of international organizations and markets on education policy making*. Basingstoke: Palgrave Macmillan, pp. 40–56.
- Meltzoff, A.N., Kuhl, P.K., Movellan, J. and Sejnowski, T.J. (2009) 'Foundations for a new science of learning', *Science*, 325(5938), pp. 284–288.
- Menon, V. (2015). 'Large-scale functional brain organization', *Brain Mapping: An Encyclopedic Reference*, 2, pp. 449–459.
- Meyer, H.D. and Benavot, A. (eds.) (2013) *PISA, power, and policy: the emergence of global educational governance*. Providence, RI: Symposium Books.
- Meyer, J.W., Kamens, D.H. and Benavot, A. (1992) *School knowledge for the masses*. London: Routledge.
- Mincer, J.A. (1974) *Schooling, experience, and earnings*. Cambridge: National Bureau of Economic Research.
- Mislevy, R.J. (1994). 'Evidence and inference in educational assessment', *Psychometrika*, 59(4), pp. 439–483.
- Mislevy, R.J. and Haertel, G.D. (2006). 'Implications of evidence-centered design for educational testing', *Educational Measurement: Issues and Practice*, 25(4), pp. 6–20.
- Mo, J. (2017) *Collaborative problem solving* (no. 78). Paris: OECD.
- Mok, K. and Jiang, J. (2017) 'Massification of higher education and challenges for graduate employment and social mobility: East Asian experiences and sociological reflections', *International Journal of Educational Development*, 63, pp. 44–51.
- Morrison, J.R., Reilly, J.M. and Ross, S.M. (2019) 'Getting along with others as an educational goal: an implementation study of Sanford', *Journal of Research in Innovative Teaching and Learning*, 12(1), pp. 16–34.
- Muller, J.Z. (2018) *The tyranny of metrics*. Princeton: Princeton University Press.
- Mundy, K. (1998) 'Educational multilateralism and world (dis)order',



Comparative Education Review, 42(4), pp. 448–478.

Mundy, K. (2006) 'Education for all and the new development compact', International Review of Education, 52(1/2), pp. 23–48.

Munroe, E., Borden, L., Murray Orr, A., Toney, D. and Meader, J. (2013) 'Decolonizing Aboriginal education in the 21st century', McGill Journal of Education, 48(2), pp. 317–337.

National Research Council (2015) Transforming the workforce for children birth through age 8: a unifying foundation. Washington, DC: The National Academies Press.

Ngugi, W.T. (1986) Decolonising the mind: the politics of language in African literature. Melton: James Currey.

Niu, W. (2007) 'Western influences on Chinese educational testing', Comparative Education, 43(1), 71–91.

Novoa, A.N. and Yariv-Mashal, T. (2003) 'Comparative research in education: a mode of governance or a historical journey?', Comparative Education, 39(4), pp. 423–438.

Obanya, P. (2007). Thinking and talking education. Ibadan: Evans Brothers.

OECD (2003) The PISA 2003 assessment framework: mathematics, reading, science and problem solving knowledge and skills. Paris: OECD.

OECD (2005) Formative assessment: improving learning in secondary classrooms. Paris: OECD.

OECD (2006) Education at a glance 2006: OECD indicators. Paris: OECD.

OECD (2007) 'The ethics and organisation of educational neuroscience', in Understanding the brain: the birth of a learning science. Paris: OECD.

OECD (2011a) 'Chair's summary', meeting of the OECD council at ministerial level Paris, 25–26 May. Available at: <http://www.oecd.org/mcm/oecdministerialcouncilmeeting2011chairsummary.html> (Accessed: 11 January 2021).

OECD (2011b) 'Framework for an OECD strategy on development', meeting of the OECD council at ministerial level Paris, 25–26 May. Available at: <http://www.oecd.org/development/48106820.pdf>. (Accessed: 9 January 2021).

OECD (2013a) PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy. Paris: OECD.

OECD (2013b) PISA 2015 draft collaborative problem solving framework. Paris: OECD.

OECD (2015) Beyond PISA 2015: a longer-term strategy of PISA. Paris: OECD.

OECD (2016a) Global competency for an inclusive world. Paris: OECD. Available at: <http://globalcitizen.nctu.edu.tw/wp-content/uploads/2016/12/2.-Global-competency-for-an-inclusive-world.pdf> (Accessed: 9 January 2021).

OECD (2016b) Draft framework of the PISA 2018 global competence assessment. Paris: OECD.

OECD (2018) PISA for development assessment and analytical framework: reading, mathematics and science. Paris: OECD.

OECD (2019) Fostering students' creativity and critical thinking: what it means in school. Paris: OECD.

Olivier, C. (1999) Let's educate, train and learn outcomes-based: a 3D experience in creativity. South Africa: NQF Based Design Book.

Olson, D.R. (2003) Psychological theory and educational reform: how school remakes mind and society. Cambridge: Cambridge University Press.

# REFERENCES

- Ozan, C. and Kincal, R.Y. (2018). 'The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills', *Educational Sciences: Theory and Practice*, 18(1), pp. 85–118.
- Parisi, G.I., Kemker, R., Part, J.L., Kanan, C. and Wermter, S. (2019) 'Continual lifelong learning with neural networks: a review', *Neural Networks*, 113, pp. 54–71.
- Patel, D. and Sandefur, J. (2020). *A Rosetta Stone for human capital*. Washington, DC: Center for Global Development. Available at: <https://www.cgdev.org/sites/default/files/rosetta-stone-human-capital.pdf>
- Pellegrino, J.W., Chudowsky, N. and Glaser, R. (2001) *Knowing what students know: the science and design of educational assessment*. Washington, DC: The National Academies Press.
- Poldrack, R.A. (2012) 'The future of fMRI in cognitive neuroscience', *Neuroimage*, 62(2), pp. 1216–1220.
- Posner, M.I. and Rothbart, M.K. (2005) 'Influencing brain networks: implications for education', *Trends in Cognitive Sciences*, 9(3), pp. 99–103.
- Price, C.J. (2012) 'A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading', *Neuroimage*, 62(2), pp. 816–847.
- Psacharopoulos, G. (1994) 'Returns to investment in education: a global update', *World Development*, 22(9), pp. 1325–1343.
- Psacharopoulos, G. and Patrinos, H.A. (2004) 'Returns to investment in education: a further update', *Education Economics*, 12(2), pp. 111–134.
- Pykett, J. (2015) *Brain culture: shaping policy through neuroscience*. Bristol: Policy Press.
- Rivera, S.M., Reiss, A.L., Eckert, M.A. and Menon, V. (2005) 'Developmental changes in mental arithmetic: evidence for increased functional specialization in the left inferior parietal cortex', *Cerebral Cortex*, 15(11), pp. 1779–1790.
- Roediger III, H.L. and Karpicke, J.D. (2006a). 'Test-enhanced learning: taking memory tests improves long-term retention', *Psychological Science*, 17(3), pp. 249–255.
- Roediger III, H.L. and Karpicke, J.D. (2006b). 'The power of testing memory: basic research and implications for educational practice', *Perspectives on Psychological Science*, 1(3), pp. 81–210.
- Rogiers, A., Merchie, E. and Keer, H.V. (2019) 'Learner profiles in secondary education: occurrence and relationship with performance and student characteristics', *The Journal of Educational Research*, 112(3), pp. 385–396.
- Ross, S.M. and Morrison, J.R. (2020). 'Achieving better educational practices through research evidence: a critical analysis and case illustration of benefits and challenges', *ECNU Review of Education*, 4(1), pp. 108–127.
- Rutkowski, L., Rutkowski, D. and Liaw, Y.L. (2019) 'The existence and impact of floor effects for low-performing PISA participants', *Assessment in Education: Principles, Policy & Practice*, 26(6), pp. 643–664.
- Sälzer, C. and Roczen, N. (2018). 'Assessing global competence in PISA 2018: challenges and approaches to capturing a complex construct' *International Journal of Development Education and Global Learning*, 10(1), pp. 5–20.
- Schiefelbein, E. and Schiefelbein, P. (2003). 'From screening to improving quality: the case of Latin America', *Assessment in Education*, 10(2), pp. 141–154.
- Schildkamp, K. and Kuiper, W. (2010). 'Data-informed curriculum reform: which data, what purposes, and promoting and hindering factors', *Teaching and Teacher Education*, 26(3), pp. 482–496.
- Schultz, T.W. (1961) 'Investment in human capital', *The American Economic Review*, 51(1), pp. 1–17.
- Schultz, T.W. (1971) *Investment in human capital: the role of education and of research*. New York: Free Press.



- Schuwirth, L.W. and Van der Vleuten, C.P. (2011) 'Programmatic assessment: from assessment of learning to assessment for learning', *Medical Teacher*, 33(6), pp. 478–485.
- Sedlacek, W.E. (1994) 'Issues in advancing diversity through assessment', *Journal of Counseling & Development*, 72, pp. 549–553.
- Seghier, M.L., Fahim, M.A. and Habak, C. (2019) 'Educational fMRI: from the lab to the classroom', *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02769>.
- Sellar, S. and Lingard, B. (2013). 'PISA and the expanding role of the OECD in global educational governance', in PISA, power, and policy: The emergence of global educational governance. Oxford: Symposium Books, pp. 185–206.
- Sellar, S. and Lingard, B. (2014). 'The OECD and the expansion of PISA: new global modes of governance in education', *British Educational Research Journal*, 40(6), pp. 917–936.
- Sen, A. (1999) *Development as freedom*. Oxford: Oxford University Press.
- Shore, R. and Bryant, J. (2011) 'A role for neuroscience in shaping contemporary education policy', *AASA Journal of Scholarship & Practice*, 8(3), pp. 30–42.
- Sigman, M., Pena, M., Goldin, A.P. and Ribeiro, S. (2014) 'Neuroscience and education: prime time to build the bridge', *Nature Neuroscience*, 17(4), pp. 497–502.
- Sjøberg, S. (2016). 'OECD, PISA and globalisation: the influence of the International Assessment Regime', in Tienken, C.H. and Mullen, C.A. (eds.) *Education policy perils: tackling the tough issues*. New York: Routledge.
- Smedt, B.D. (2018) 'Applications of cognitive neuroscience in educational research', *Oxford Research Encyclopedia of Education*. doi: <https://doi.org/10.1093/acrefore/9780190264093.013.69>.
- Smeyers, P. (2016) 'Neurophilia: guiding educational research and the educational field?', *Journal of Philosophy of Education*, 50(1), pp. 62–75.
- Smith, C., Parr, N. and Muhidin, S. (2019) 'Mapping schools' NAPLAN results: a spatial inequality of school outcomes in Australia', *Geographical Research*, 57(2), pp. 133–150.
- Solano-Flores, G. and Nelson-Barber, S. (2001) 'On the cultural validity of science assessments', *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 38(5), pp. 553–573.
- Sousa, D.A. (2006) *How the brain learns*. Thousand Oaks, CA: Corwin Press.
- Spiller, D. (2012) *Assessment matters: self-assessment and peer assessment*. Hamilton: The University of Waikato.
- State Council (2014) *State Council directive on deepening the implementation of the reform of the examination enrollment system*. Available at: [http://www.gov.cn/zhengce/content/2014-09/04/content\\_9065.htm](http://www.gov.cn/zhengce/content/2014-09/04/content_9065.htm) (Accessed: 30 January 2021).
- Sternberg, R.J. and Grigorenko, E.L. (2001) 'All testing is dynamic testing', *Issues in Education*, 7(2), pp. 137–170.
- Sternberg, R.J. and Grigorenko, E.L. (2002) *Dynamic testing: the nature and measurement of learning potential*. Cambridge: Cambridge University Press.
- Stoffer, J. (2017) 'The importance of culturally safe assessment tools for Inuit students', *The Australian Journal of Indigenous Education*, 46(1), pp. 64–70.
- Sylwester, R. (1995) *A celebration of neurons: an educator's guide to the human brain*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Szucs, D. and Ioannidis, J.P.A. (2017) 'Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature', *PLOS Biology*, 15(3). doi: <https://doi.org/10.1371/journal.pbio.2000797>.
- Szucs, D. and Ioannidis, J.P.A. (2020) 'Sample size evolution in neuroimaging research: an evaluation of highly-cited

# REFERENCES

- studies (1990–2012) and of latest practices (2017–2018) in high-impact journals', *NeuroImage*, 221. doi: 10.1016/j.neuroimage.2020.117164.
- Tabb, W.K. (2006) Microsoft encarta online encyclopedia 2006. Available at: [http://encarta.msn.com/text\\_1741588397\\_1/Globalization.html](http://encarta.msn.com/text_1741588397_1/Globalization.html) (Accessed: 30 January 2021)
- Tan, C. and Ng, C. (2018) 'Assessment reform in Shanghai: issues and challenges', *International Journal of Educational Reform*, 27(3), pp. 291 – 309.
- Taras, M. (2010) 'Student self-assessment: processes and consequences', *Teaching in Higher Education*, 15(2), pp. 199–209.
- Thomas, M.S.C., Ansari, D. and Knowland, V.C.P. (2019) 'Annual research review: educational neuroscience: progress and prospects', *Journal of Child Psychology and Psychiatry*, 60(4), pp. 477–492.
- Tikly, L. (2017). 'The future of education for all as a global regime of educational governance', *Comparative Education Review*, 61(1), pp. 22–57.
- Turner, B.O., Paul, E.J., Miller, M.B and Barbey, A.K. (2018) 'Small sample sizes reduce the replicability of task-based fMRI studies', *Communications Biology*, 1(1), pp. 1–10.
- UN (2014) *The road to dignity by 2030: ending poverty, transforming all lives and protecting the planet: synthesis report of the Secretary-General on the post-2015 sustainable development agenda*. New York: United Nations.
- UN (2015) *Transforming our world: the 2030 agenda for sustainable development*. New York: United Nations.
- UNESCO (2013) *Global citizenship education: preparing learners for the challenges of the twenty-first century*. Paris: UNESCO.
- UNESCO (2015) *Rethinking education: towards a global common good?* Paris: UNESCO. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000232555> (Accessed: 5 January 2022).
- UNESCO IITE (2020) *AI in education: change at the speed of learning*. Available at: [https://iite.unesco.org/wp-content/uploads/2020/11/Steven\\_Duggan\\_AI-in-Education\\_2020.pdf](https://iite.unesco.org/wp-content/uploads/2020/11/Steven_Duggan_AI-in-Education_2020.pdf) (Accessed: 1 February 2021).
- UNESCO MGIEP (2017) *Rethinking schooling for the 21st century: the state of education for peace, sustainable development and global citizenship in Asia*. New Delhi: UNESCO MGIEP. Available at: [https://d1c337161ud3pr.cloudfront.net/files%2F20d85fcb-1b92-476e-9749-c23f0d794d48\\_Rethinking%20Schooling%20for%20the%2021st%20Century.pdf](https://d1c337161ud3pr.cloudfront.net/files%2F20d85fcb-1b92-476e-9749-c23f0d794d48_Rethinking%20Schooling%20for%20the%2021st%20Century.pdf). (Accessed: 3 March 2020).
- Unterhalter, E. (2009). 'Social justice, development theory and the question of education', in Cowen, R. and Kazamias, A.M. (eds.) *International handbook of comparative education*. Dordrecht: Springer Netherlands, pp. 781–800.
- Unterhalter, E. (2016). 'Negative capability? Measuring the unmeasurable in education', *Comparative Education*, 53(1), pp. 1–16.
- Vaccari, V. and Gardinier, M.P. (2019) 'Toward one world or many? A comparative analysis of OECD and UNESCO global education policy documents', *International Journal of Development Education and Global Learning*, 11(1), pp. 68–86.
- Varma, S., McCandliss, B.D. and Schwartz, D.L. (2008) 'Scientific and pragmatic challenges for bridging education and neuroscience', *Educational Researcher*, 37(3), pp. 140–152.
- Vickers, E. and Zeng, X.D. (2017) *Education and society in post-Mao China*. Abingdon: Routledge.
- Vogeley, K. and Roepstorff, A. (2009). 'Contextualising culture and social cognition', *Trends in Cognitive Sciences*, 13(12), pp. 511–516.
- Wan, Y. and Vickers, E. (2021) 'Towards meritocratic apartheid? Points systems and migrant access to China's urban public schools'. *The China Quarterly*, 1-29. doi:10.1017/S0305741021000990.

- Wang, Y. and Ross, H. (2010) 'Experiencing the change and continuity of the College Entrance Examination' *Chinese Education & Society*, 43(4), pp. 75–93.
- Watts, N., Amann, M., Arnell, N., Ayeb-Karlsson, S., Belesova, K., ... (2019). 'The 2019 report of The Lancet Countdown on health and climate change: ensuring that the health of a child born today is not defined by a changing climate' *The Lancet*, 394(10211), pp. 1836–1878.
- WHO (2018) *Survive and thrive: transforming care for every small and sick newborn: key findings* (No. WHO/FWC/MCA/18.11). Geneva: World Health Organization.
- William, D. (2011) 'What is assessment for learning?', *Studies in Educational Evaluation*, 37(1), pp. 3–14.
- Williamson, B. (2017) *Big Data in education: the digital future of learning, policy and practice*. Thousand Oaks, CA.
- Wiseman, A.W. (2010) 'The uses of evidence for educational policymaking: global contexts and international trends', *Review of Research in Education*, 34(1), pp. 1–24.
- Wiseman, A.W. and Chase-Mayoral, A. (2014) 'Shifting the discourse on neo-institutional theory in comparative and international education', in *Annual review of comparative and international education*. Bingley: Emerald Group Publishing Limited, pp. 99–126.
- Wolfe, P. and Brandt, R. (1998) 'What do we know from brain research?', *Educational Leadership*, 56(3), pp. 8–13.
- Woodward, R. (2004) 'The Organisation for Economic Co-operation and Development: global monitor', *New Political Economy*, 9(1), pp. 113–127.
- Wu, X., Wu, R., Chang, H.H., Kong, Q. and Zhang, Y. (2020) 'International comparative study on PISA mathematics achievement test based on cognitive diagnostic models', *Frontiers in Psychology*, 11. doi: <https://doi.org/10.3389/fpsyg.2020.02230>.
- Yarkoni, T. (2009) 'Big correlations in little studies: inflated fMRI correlations reflect low statistical power: commentary on Vul et al. (2009)', *Perspectives on Psychological Science*, 4(3), pp. 294–298.
- Yasukawa, K., Hamilton, M. and Evans, J. (2017) 'A comparative analysis of national media responses to the OECD survey of adult skills: policy making from the global to the local?', *Compare: A Journal of Comparative and International Education*, 47(2), pp. 271–285.
- Ydesen, C. and Grek, S. (2020) 'Securing organisational survival: a historical inquiry into the OECD's work in education during the 1960s', *Paedagogica Historica*, 56(3), pp. 412–427.
- You, Y. and Morris, P. (2016) 'Imagining school autonomy in high-performing education systems: East Asia as a source of policy referencing in England', *Compare: A Journal of Comparative and International Education*, 46(6), pp. 882–905.
- Zhen, Z. (2017) 'Strive to build a modern examination enrollment system with Chinese characteristics'. *China Examinations*, 5(301), pp. 1–5.
- Zhou, J. and Fischer, K.W. (2013). 'Culturally appropriate education: insights from educational neuroscience', *Mind, Brain, and Education*, 7(4), pp. 225–231.
- Zins, J.E. and Elias, M.J. (2007) 'Social and emotional learning: promoting the development of all students', *Journal of Educational and Psychological Consultation*, 17(2–3), pp. 233–255.