

## EXTRACTS FROM: “ RETHINKING ASSESSMENT IN EDUCATION: THE CASE FOR CHANGE”, CSE LEADING EDUCATION SERIES, BILL LUCAS, APRIL 2021

Across the world assessment is not working. We are not evidencing the kinds of dispositions and capabilities that society increasingly wants. Educational jurisdictions are placing too much reliance on high-stakes, standardised testing. They are testing the wrong things in the wrong ways. High-stakes assessment is having a damaging impact on the health and wellbeing of students and it is not giving universities, colleges or employers the kind of information they want.

Assessment is out of sync with curriculum and pedagogy. Where we have become increasingly evidence-based in teaching and learning, we are failing to keep up with the science of assessment, preferring to rely on outdated, outmoded and unsubtle methods.

Our young people require all of us working in education to establish greater clarity about the uses of assessment in education, linked to a greater understanding of the science of assessment.

We need nothing less than a paradigm shift in our understanding about how best to create assessment systems that use more effective ways of evidencing the full range of student progress.

### The wrong kind of nets for catching young people’s strengths

*To solely use standardised achievement tests is like casting a net into the sea – a net that is intentionally designed to let the most interesting fish get away. Then, to describe the ones that are caught strictly in terms of their weight and length is to radically reduce what we know about them. To further conclude that all the contents of the sea consist of fish like those in the net compounds the error further. We need more kinds of fish. We need to know more about those we catch. We need new nets. (William T Randolph, Commissioner of Education, Colorado1)*

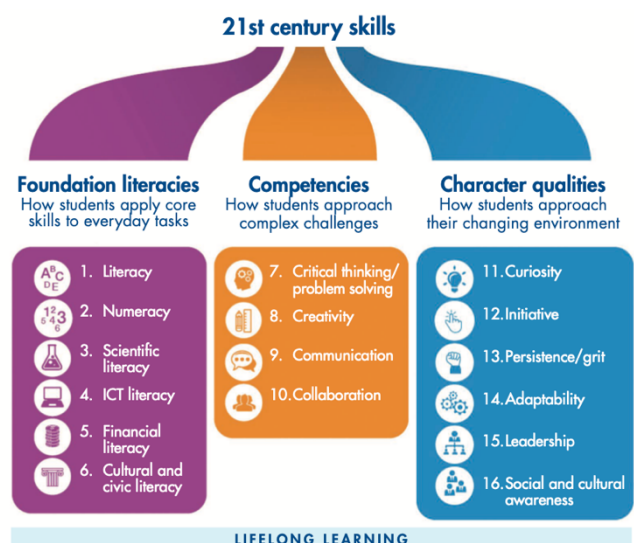
When it comes to assessment, William Randolph’s thoughtlessly designed net seems an apt image for our times. For, in different ways, educational assessment systems across the world have become very good at weighing and measuring students, without reflecting on whether the assessments they are making are relevant, meaningful or useful, and without considering the consequences of the assessment process.

### The Beginnings of a Global Curriculum

Today there are a dozen or so well-regarded models of what contemporary curricula should look like. The World Economic Forum (2015) is widely cited (Figure 1). Whether framed as foundational literacies, competencies or character qualities, it is increasingly recognised that, in our digital age, there are more core literacies than we once thought.

In such a view of education we might be asking questions such as:

- *What kind of knowledge is it important for all young people to have?*
- *What kind of dispositions is it important for all young people to acquire?*
- *How can we ensure that young people acquire and apply useful knowledge in a range of settings?*
- *How can we teach young people to work across subject disciplines, as happens in the real world, ensuring that they have the necessary building blocks in place?*



- *How can we ensure that important dispositions for learning and for life are best cultivated in a range of disciplinary contexts?*
- *How can we develop strength, breadth and depth in learning to facilitate its transfer across contexts?*
- *Which pedagogies work best for promoting deep learning?*
- *How best to assess knowledge and evidence dispositions?*

## Learning 2.0

Richard Elmore (2019) helpfully summarises the consequences for school of two very different conceptions of learning, as follows.

### *Learning 1:*

Learning is the ability to recall and deploy information and algorithms accurately and appropriately.

Schooling is the mechanism by which we organise social and status consistent with this definition of learning.

Assessment is the means by which we define, measure, evaluate, and confer 'merit', consistent with this definition of learning.



### *Learning 2:*

Learning is the ability to consciously modify understandings, beliefs and actions in response to evidence, experience, and reflection.

Schooling is one of many environments in which humans develop the capability to exercise judgement and control over what they learn, how they learn, and what they intend to do with what they have learned.

Assessment is the means by which individuals receive useful information about the development of their capabilities as learners over time. (Elmore, 2019, p 333)

In this paper it is the second of these two conceptions that we shall be exploring.

## The tail that wags the dog

*Assessment influences not just what gets taught but how it gets taught.*

Almost anyone who has worked in education knows that what gets assessed by and large gets taught. You can have a bold and expansive curriculum, but as the time of examinations draws close, the focus shifts to those aspects of the curriculum which will be assessed. The decisions are complex for young people as they navigate their next steps, and the means by which such decisions are made are often by 'high- stakes assessment'.

Assessment influences not just what gets taught but how it gets taught. If Teacher X uses a particular teaching method for science with her class and students do well in their assessments, while Teacher Y uses a different method and her class does less well, then, assuming the classes share similar enough characteristics, schools and school systems will begin to draw lessons from this. Reasonably enough they will suggest that when teaching science the methods chosen by Teacher X are the ones to use.

At first sight this is an intelligent system's response, but what if assessments in science privilege decontextualised recall of scientific theory and simplistic memorisation of scientific facts, which neither encourage students to think and work like scientists nor equip them to go on to deeper study of science and its uses in society? What if thinking about assessment is not keeping up with advances in the learning sciences? In these cases such a response would be dumb.

The dog in the sub-heading of this section is the school system and the tail that wags it is, of course, assessment. Also, mixing my metaphors, the tails which seem to wag so many school systems across the world are the fishing nets with which this section began.

Eight years ago Geoff Masters suggested that the 'field of educational assessment is currently divided and in disarray' in Australia (Masters, 2013, p 1). I suspect that this is still the case in Australia and still the case for the majority of educational jurisdictions across the world today.

Let us turn now to what is wrong with educational assessment in more detail.

## THE PROBLEM WITH EDUCATIONAL ASSESSMENT TODAY

*The measurement of deep learning must be always informed by a wealth of underlying assessment evidence that captures the complete picture of who students are, what they know and whether they are prepared to use that knowledge to advance their lives and others. (Joanne McEachen, Assessment for Deep Learning, 2017, p 12)*

There are many aspects of educational assessment today which are failing. These fall into the four broad areas of

1. what is assessed (focus);
2. how it is assessed (methods);
3. the impact of the assessment process (consequences); and
4. the uses made of the assessment (validity).

Of course, there is also a fifth challenge: the degree to which whatever we might want to measure can be reliably assessed. In a recent review (2020a) Sandra Milligan and colleagues cut across all of these categories elegantly when they suggested that:

Without a focus on mastery of generic capabilities, assessment and teaching practices tend to privilege memorisation, essay writing, individual mastery of set content and solving of problems with formulaic solutions. The risk is that schools create students dependent on direct instruction, cramming, drilling and coaching, reliant on expert instruction by teachers who are expected to guide learners through a carefully prescribed body of knowledge, assessed in predictable ways. (p 14)

### An assessment focus that is too shallow and too narrow

Currently, the knowledge that is typically assessed is from a narrow range of subjects, rarely explored in depth and almost never interdisciplinary. Practical knowledge and skill is not much assessed in general education, and individuals rather than teams remain the focus.

Assessments frequently require recall of content but rarely demand the kind of deep thinking, problem solving or application needed in the real world.

Complex, higher order skills are rarely assessed in ways that recognise the subtleties involved. Many dispositions or capabilities known to be important in life are not assessed at all. – Darling-Hammond, 2017

- Traditional areas, literacy, maths and science continue to require considerable content to be tested, while newer areas such as citizenship, sustainable development and ethical understanding are only briefly explored.
- Except in a very few countries (Finland and Singapore are examples) there is little or no interdisciplinary assessment.
- Practical knowledge and skill is rarely assessed even in those subjects where it once used to be a central component, such as science.
- Students' capabilities in planning and undertaking extended investigations are rarely assessed.
- Although the ability to collaborate with others is widely valued in the workplace it is only acknowledged at school on the sports field or in music and drama performances.

- While dispositions or capabilities are becoming more visible in curricula they are rarely assessed; at a global level PISA's innovative domain tests of collaborative problem-solving and creative thinking are exceptions, as is the State of Victoria's testing of critical and creative thinking.

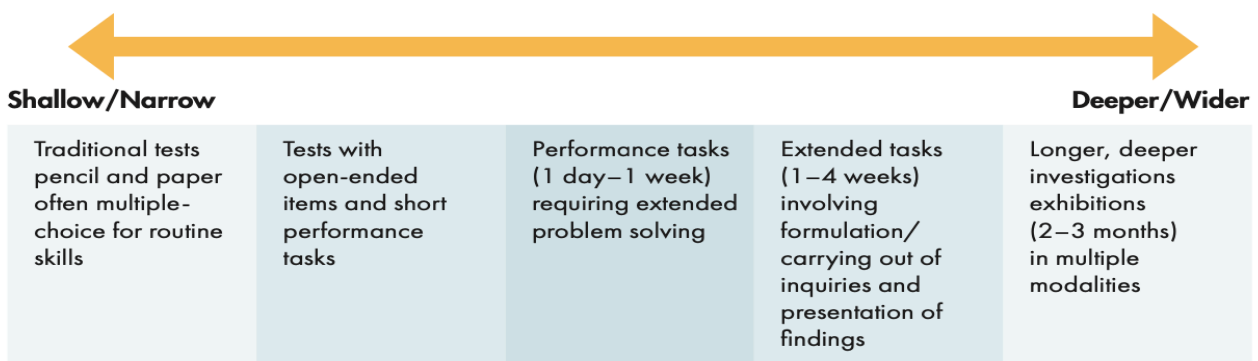
**Assessment methods that are too blunt and too narrow**

Most tests used in schools still rely on paper and pencil. They examine aspects of knowledge and routine skills. They test students' ability to remember and write about something, rather than apply or do the thing they have been learning. Concepts and skills are tested in individual subjects and only very rarely across disciplines.

While tests often purport to be criterion-based, many countries effectively revert to norm-referencing either because of the scale used (the ATAR in Australia, for example). Even where tests are explicitly criterion-based, grades often relate to syllabus content, rather than to more carefully sequenced learning progressions.

Traditional assessment methods typically fail to measure the high-level skills, knowledge, attributes and characteristics of self-directed and collaborative learning that are increasingly important for our global economy and fast-changing world. (Griffin, McGaw and Care, 2012, p v-vi).

**Figure 3. A continuum of assessment methods, adapted from Darling-Hammond (2017), p 6**



A recent High Resolves report (2020) proposes the concept of 'strings-based assessment' to exemplify the kind of blend or 'strings' of immersive, repeated practices and real-world applications that may be useful in evidencing high-order skills in citizenship education.

The range of possible assessment methods educational jurisdictions might choose from is actually wide (see the continuum of assessment methods shown in figure 3).

*Fundamentally, most assessments fail to capture the degree to which students have progressed over time.*

Students are tested at set times rather than when they are ready, often to meet the needs of the next educational provider. These inflexible encounters with assessment ignore the huge variety of student achievement levels, where 'in any given year of school, the most advanced learners in areas such as Reading and Mathematics can be as much as five or six years ahead of the least advanced learners' (Masters, 2013, p 3), the fact that 'attainment is only loosely related to age' (William, 2007) and the differing levels of maturity found in any cohort on account of birth dates.

More fundamentally, most assessments fail to capture the degree to which students have progressed over time. Instead they:

*... provide snapshots of achievement at particular points in time, but they do not capture the progression of students' conceptual understanding over time, which is at the heart of learning. This limitation exists largely because most current modes of assessment lack an underlying theoretical framework of how student understanding in a content domain develops.* (Pellegrino, Chudowsky and Glaser, 2001, p 27–28).

Assessments need not be done in this way, as 'Measuring progress provides a deliberate counterpoint to the traditional practice of measuring achievement at specific time points'.

## Consequences that are unhelpful

In any assessment system there are intended and unintended consequences, but it would seem fundamental to assume that an essential principle should be, as the USA's Gordon Commission on assessment in 2013 noted, that assessment systems should 'do no harm'.

Sadly, the consequences of the focus and methods of many, especially high-stakes assessments, are well-documented and harmful in a number of ways, including:

- *leading students to conclude that they are failures (Education Policy Institute, 2019);*
- *demotivating students to the extent that they may not stay on at school or find employment (Milligan et al, 2020a);*
- *making it less likely that students will see themselves as learners and want to continue learning throughout their lives (Tuckett and Field, 2016)*
- *causing negative impact on young people's wellbeing (Howard, 2020);*
- *exacerbating inequity (Au, 2016);*
- *reducing performance through anxiety, especially for students of lower ability (von der Embse et al, 2018);*
- *distracting from the huge importance of assessment for learning and assessment as learning (Birenbaum, 2015);*
- *misunderstanding and undervaluing wider skills and dispositions by not measuring them (Heckman and Kautz, 2013), and perpetuating the myth that soft skills are easy to acquire and of less value than so-called hard skills such as core literacies;*
- *inviting a lack of trust in teacher judgement in some jurisdictions (Harlen, 2005; Coe et al, 2020) which, in an unhelpfully reinforcing loop, can lead to lower levels of teacher assessment 'literacy'.*

In *The Testing Charade* (2015), Koretz reminds us of the danger of Campbell's law, that:

*the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor (p 38) ... When test scores become the goal of the teaching process, they both lose their value as indicators of educational status and distort the educational process in undesirable ways.*

The National Academy of Education (2021) points out that, to avoid unintended and sometimes unfair consequences, we need to *Communicate clearly (and often) the intended purposes and uses of particular assessments as well as any relevant context.*

## Characteristics of high-quality assessment systems

In the last decade a significant number of reviews (Lai and Viering, 2012; Conley and Darling-Hammond, 2013; Bennett, 2013; OECD, 2013; Masters, 2013; Soland, Hamilton and Stecher, 2013; Hill and Barber, 2014; Siarova, Sternadel and Mašidlauskaitė, 2017; Care et al, 2018; O'Connell, Milligan and Bentley, 2019; Care, Anderson and Kim, 2019; Milligan et al, 2020b) have looked at the implications for systems wanting to move towards the assessment of deeper learning, what Elmore calls 'Learning 2'.

While analysing implications at a system level is complex, and needs to take into account the differences between the political intentions of educational jurisdictions, the convergence of thinking across these reviews – combined with the slowness with which their suggestions have been taken up – reminds us of how difficult it is to change assessment systems.

Common themes from evidence on high-quality assessment systems include the following.

### *Purpose and consequence*

- *The importance of understanding the purpose any assessment is intended to serve.*
- *A growing recognition of assessment as a tool for improvement at individual, school and system level.*

- *The tensions that exist between summative and formative approaches.*
- *The many unhelpful consequences of high-stakes assessment.*

### *Depth and breadth*

- *A need to evidence high-order thinking skills reliably.*
- *A requirement for better definitions of dispositions and associated learning progressions.*
- *The growing visibility of dispositions in the curricula of educational jurisdictions.*
- *The desirability of assessments being pedagogically sensitive and educationally valuable.*
- *The complexity of designing ways of fairly evidencing student progress within dispositions.*
- *A growing interest in the concept of mastery.*
- *The need for flexibility to ensure that the full range of abilities can be fairly assessed.*

### *Authenticity*

- *Increasing interest in strengths-based approaches.*
- *The need to design better performance-based assessments.*
- *A move towards assessments of investigations over longer time periods.*
- *Some interest in assessment on demand.*
- *Increased opportunities for student involvement and agency in the process.*

### *Progression and improvement*

- *The benefits of assessment for and as learning.*
- *The need for multimodal approaches to assessment, incorporating data from a number of sources.*

### *Quality infrastructure*

- *A better understanding of when to use assessment of, for and as learning.*
- *The need for new assessment partnerships.*
- *Enhanced teacher capacity in assessment literacy and moderation.*