

Rethinking assessment in education: The case for change

BILL LUCAS

Acknowledgements

All my colleagues at Rethinking Assessment, especially Rosie Clayton, Peter Hyman, Rachel Macfarlane and Al McConville; to the Edge Foundation for its support; and to the team at UCL School of Management for their research.

Many along the way whose work has influenced my thinking about assessment, including: Michelle Anderson, Ken Baker, Geoff Barton, Ron Berger, Sarah-Jayne Blakemore, Roy Blatchford, Esther Care, Christine Cawsay, Guy Claxton, Paul Collard, Art Costa, Angela Duckworth, Carol Dweck, Charles Fadel, Sharon Foster, Michael Fullan, Valerie Hannon, John Hattie, Lois Hetland, Rosemary Hipkins, David Howes, Tony Mackay, Geoff Masters, Jonnie Noakes, James Pellegrino, David Perkins, Mario Piantanini, Sandra Milligan, Andreas Schleicher, Ellen Spencer, Michael Stevenson, Louise Stoll, Stéphan Vincent-Lancrin, Dylan Wiliam.

Contents

- 2 Introduction
- 3 The wrong kind of nets for catching young people's strengths
- 5 An education system fit for purpose?
- 12 The problem with educational assessment today
- 19 Revisiting the purposes of assessment
- 23 Promising practices from across the world
- 33 Visible progress
- 38 References



ISSN 1838-8566 ISBN 978-1-925654-57-8

© 2021 Centre for Strategic Education, Victoria

The Centre for Strategic Education welcomes usage of this publication within the restraints imposed by the Copyright Act. Where the material is to be sold for profit then written authority must be obtained first. Detailed requests for usage not specifically permitted by the Copyright Act should be submitted in writing to:

The Centre for Strategic Education
Mercer House, 82 Jolimont Street,
East Melbourne VIC 3002

Produced in Australia by Centre for Strategic Education
Mercer House, 82 Jolimont Street, East Melbourne VIC 3002

Editorial Team: Tony Mackay, Keith Redman,
Murray Cropley, Andrew Miller

Rethinking assessment in education: The case for change

BILL LUCAS



Introduction

This paper is the second in the CSE Leading Education Series and a collaboration between CSE in Australia and Rethinking Assessment in England.

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.

In addition, we want to move rapidly from theoretical debate to practical prototyping and implementation.

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)

Metaphors abound in education. From the Greeks via the Romans we took the idea that a child's mind was a tabula rasa or blank slate. Children, the comparison suggests, know nothing and bring nothing; all is dependent on the experiences that adults offer them. They are empty vessels waiting for those more knowledgeable than them to fill up their minds.

ATAR is a ladder in an educational game of snakes and ladders, whose higher rungs hold out a promise of success, which turns out to equate to abstract rather than to real-world intelligence.

There are many other metaphors we might draw on that are more cheerfully expansive: a search for hidden treasure; an odyssey; discovery; challenge.

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.

In the UK, for example, GCSE exams routinely fail 33 per cent of all sixteen year olds. The students who do not make the grade have become known as the 'forgotten third' (Association of School and College Leaders, 2019). The system has sifted 'sheep' from 'goats', but the public, the shepherd in this analogy, has little understanding of what it all means and the goats, the third who 'fail', are left with nothing much to show for their compulsory schooling.

In Australia, the Australian Tertiary Admission Rank (ATAR) is a kind of Randolph net, too. The score out of 100 gives Australian youth a certain kind of weight and length and then produces a rank order. ATAR is a ladder in an educational game of snakes and ladders, whose higher rungs hold out a promise of success, which turns out to equate to abstract rather than to real-world intelligence.

The Randolph net metaphor originated in the USA, where, notwithstanding considerable varieties in provision between states, it is reasonable to assume that it has currency as a provocation beyond Colorado.

Words matter too. The nouns and verbs we use in connection with assessment come freighted with semantic and educational baggage.

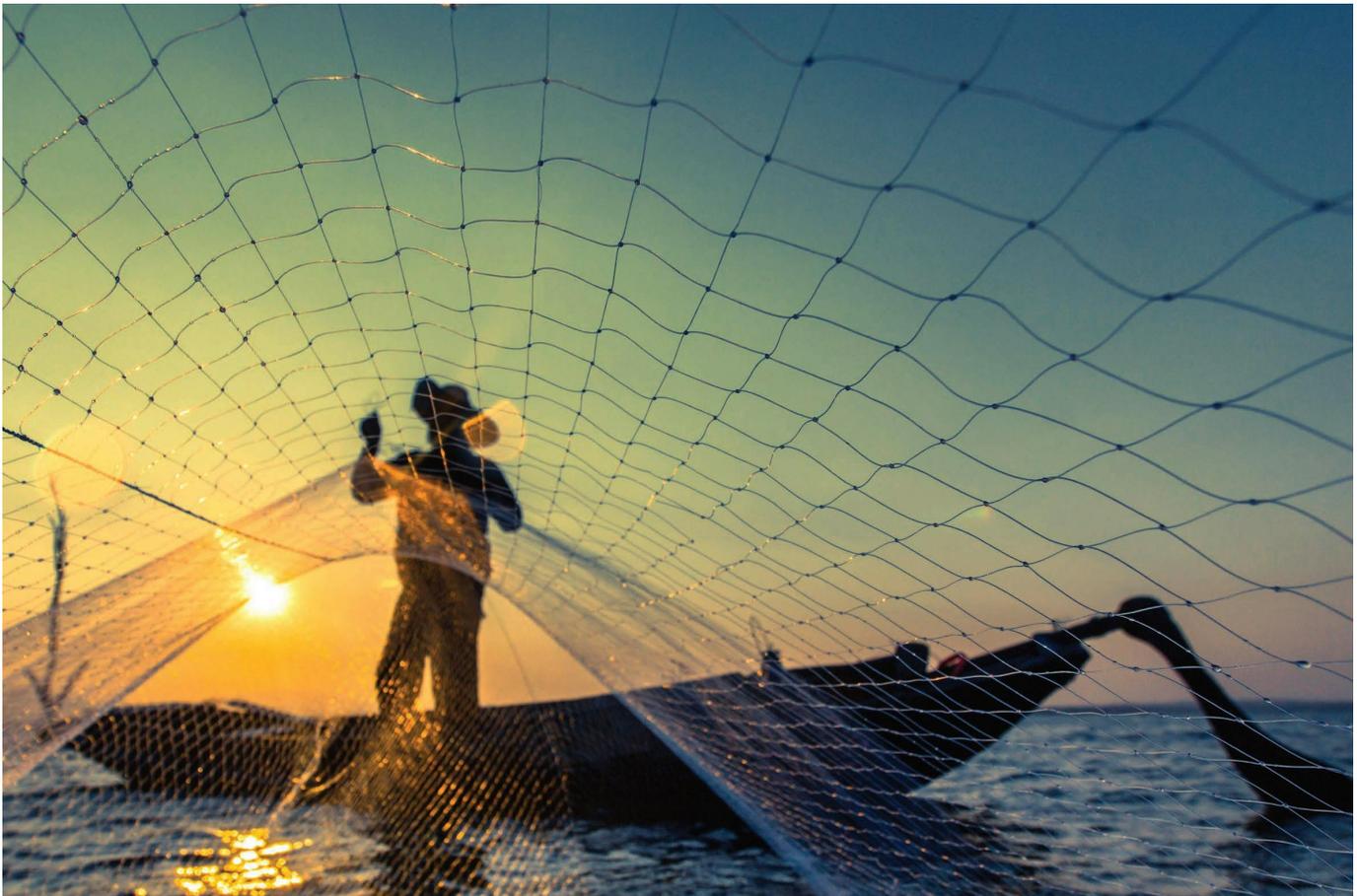
achievement, attainment, assessment, baccalaureate, badge, (balanced) score-card, competition, curriculum vitae, demonstration, diploma, evidence, exam, exhibition, expedition, feedback, illustration, interview, observation, passport, performance, portfolio, presentation, profile, project, publication, qualification, record, score, score-card, task, test, transcript, viva ...

achieve, attain, assess, curate, demonstrate, display, evidence, examine, exhibit, illustrate, measure, present, qualify, record, score, test, track (the progress of) ...

Each of these words says something about the kind of learning imagined, the method by which it might be assessed, the perspective from which such assessment is made and the validity or value that might be placed on the approach being described.

As we unpack the practices of assessment, it may be helpful to stay close to the words which have least baggage, such as 'evidence' or 'record' (noun and verb) and phrases like 'track the progress of'. Too often we invest the scores and grades used in end of school qualifications with a scientific validity they do not possess.

We need new educational nets to catch more of the capabilities young people need to thrive today.



An education system fit for purpose?

The pervasive obsession with academic grades and degrees, and corresponding elite rewards at the expense of other people ... results in narrow learning that severely distorts what people learn and need in the 21st century.

Michael Fullan, 2021, p 8

Across the world there has been growing discontent with the content of school curricula, ever since the arrival of the twenty-first century with its attendant millennial symbolism. In addition to traditional subjects such as literacy, maths and science, it is widely argued that schools need to focus on what students can **do** and **who** they are becoming.

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 (see 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

Figure 1. 16 skills for the twenty-first century (World Economic Forum, 2015)

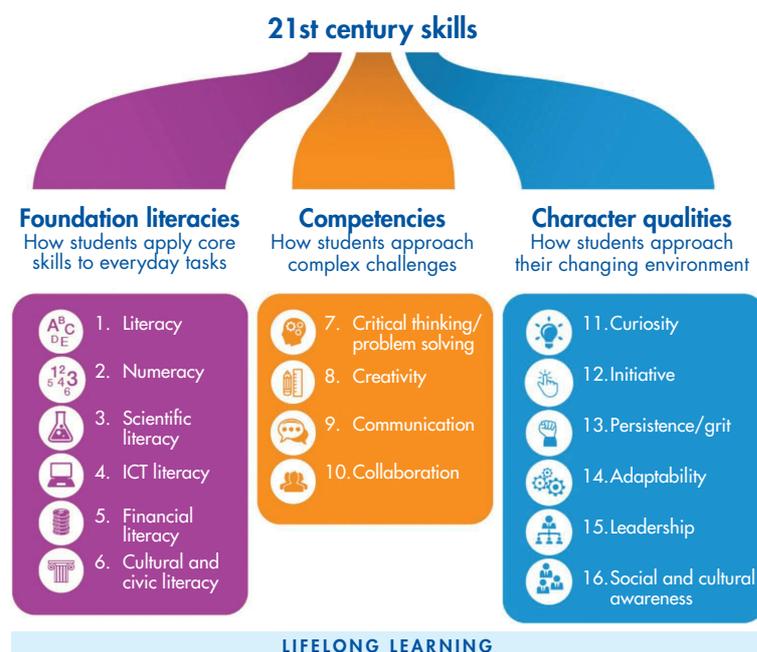
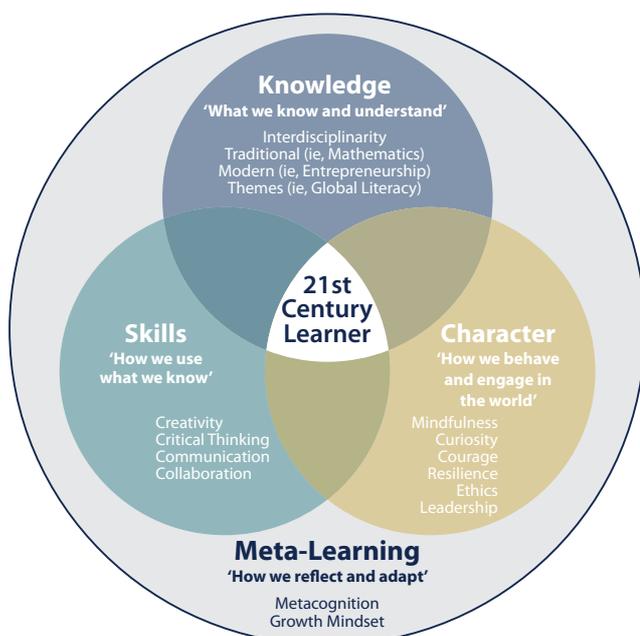


Figure 2. Center for Curriculum Redesign 4D Framework 1.0²



© Centre for Curriculum Redesign

we once thought. Whether we use words like ‘competencies’ or ‘character’ (or both) there are certain important dispositions or capabilities for living a good life, and for being a good learner, which schools have a role in cultivating.

Another model with a broadly similar framing of a contemporary curriculum

is the one developed by the Center for Curriculum Redesign (see Figure 2).

There are many variants of such contemporary curricula – of which the one Guy Claxton and I developed, *Educating Ruby: What Our Children Really Need to Learn* (2015), focusing on Confidence, Curiosity, Collaboration, Communication, Creativity, Commitment and Craftmanship is one. The approach adopted in *New Pedagogies for Deeper Learning*³ (Character, Citizenship, Collaboration, Communication, Creativity and Critical Thinking) is another. We are building on the 3Rs of old to develop the 6 or 7Cs of today.

Whether people like or do not like the framing of these dispositions as twenty-first century skills, or students as twenty-first century learners (I do not, finding the phrases vague, misleading and somewhat evangelical), there is now substantial common ground as to what these dispositions or wider skills are. Importantly this consensus includes perspectives from educationalists, parents, psychologists and researchers, as well as employers.

Table 1 is an overview of different evidence-based lists of such dispositions (Lucas, 2019).

Table 1. Dispositions for a lifetime of learning (Lucas, 2019)

European Key Competences for Lifelong Learning, 2007	Pellegrino and Hilton, 2012	Gutman and Schoon, 2013	Heckman and Kautz, 2013	Lamb et al, 2017
<ul style="list-style-type: none"> • Communication in mother tongue • Communication in foreign languages • Digital competence • Learning to learn • Social and civic competences • Sense of initiative and entrepreneurship • Cultural awareness and expression 	<ul style="list-style-type: none"> • Critical thinking • Information literacy • Reasoning • Innovation • Intellectual openness • Work ethic • Conscientiousness • Positivity • Communication • Collaboration • Responsibility • Conflict resolution 	<ul style="list-style-type: none"> • Motivation • Perseverance • Self-control • Metacognitive strategies • Social competencies • Resilience and coping • Creativity 	<ul style="list-style-type: none"> • Perseverance • Self-control • Trust • Attentiveness • Self-esteem and self-efficacy • Resilience to adversity • Openness to experience • Empathy • Humility • Tolerance of diverse opinions • Engaging productively in society 	<ul style="list-style-type: none"> • Critical thinking • Creativity • Metacognition • Problem-solving • Collaboration • Motivation • Self-efficacy • Conscientiousness • Perseverance

A generic term, expansive education, has been developed by the author and colleagues at the University of Winchester, (Lucas, Claxton and Spencer, 2013) to describe the kinds of dispositions that are desirable for success at school and in life, and how these can be interleaved into the subject disciplines of the school timetable.

Research by the Brookings Institution (Care et al, 2016) has shown that, across the world, such dispositions are gradually beginning to filter their way into schools, with 36 countries mentioning them explicitly, 76 countries identifying skills related to them, 51 locating them within the curriculum and 11 mapping their progression over the lifetime of formal schooling. The scope and sequence documents of the Australian capabilities are an example of the last of these categories. Recently the Brookings Institution, using the Center for Curriculum Redesign model, has produced an overview of the prevalence of dispositions/competencies in different educational jurisdictions (see Table 2).

Of course curriculum is only one of the three core elements of education systems, the other two being pedagogy or instruction (how the curriculum is taught) and assessment (how performance and progress is evidenced). Of significance for this report is the chronic disjoint between curriculum, pedagogy and assessment with, as yet, no focus or guidance on the teaching or assessment of these competencies/dispositions.

Interestingly, the same Brookings research shows a kind of league table of progress in developing more expansive curricula in countries and states across the world (see Table 3).

While the research does not cover every educational jurisdiction in the world, it is worth noting that the countries and states making most progress in implementing new thinking about contemporary curricula according to this report are Australia, British Columbia (Canada), Singapore, Finland, Hong Kong, Victoria (Australia) and New Zealand.

Table 2. The prevalence of Center for Curriculum Redesign competencies, (Taylor et al, 2020)

	Competency	Inclusion	Identification	Progression	Pedagogy	Assessment
Skills	Creativity	21	12	5	0	0
	Critical thinking	21	11	6	0	0
	Communication	22	11	5	0	0
	Collaboration	21	10	6	0	0
Character	Mindfulness	17	10	5	0	0
	Curiosity	17	7	3	0	0
	Courage	9	5	5	0	0
	Resilience	15	8	6	0	0
	Ethics	18	10	4	0	0
	Leadership	10	7	4	0	0
Meta-learning	Metacognition	14	7	5	0	0
	Growth mindset	14	6	5	0	0

Table 3. The frequency of competencies/dispositions across jurisdictions (Taylor et al, 2020)

Jurisdiction	CRE	CRI	COM	COL	MIN	CUR	COU	RES	ETH	LEA	MET	GRO	Total
Australia (Federal)	3	3	3	3	3	3	3	3	3	3	3	3	36
British Columbia (Canada)	3	3	3	3	3	3	3	3	3	3	3	3	36
Singapore	3	3	3	3	2	3	3	3	3	2	2	3	33
Finland	3	3	3	3	3		3	3	2	3	3	3	32
Hong Kong (China SAR)	2	3	3	3	3	1	3	3	2	2	3	3	31
Victoria (Australia)	3	3	1	3	3	1		3	3	3	3		26
New Zealand	2	2	2	2	2	2		2	2	2	2	2	22
Portugal	2	2	2	2	2	2		2	2				16
Chinese Taipei (aka Taiwan)	2	2	2	2	2	2			2				14
Denmark	2	1	1	1	1	1	1		1	1	1		11
England (UK)	2	1	1	1		2	1		1		1	1	11
Scotland (UK)	1	1	1	1	1	1	1	1	1		1	1	11
South Korea	2	2	2	1	2	1			1				11
Alberta (Canada)	1	1	1	1	1	1		1		1	1	1	10
New Brunswick (Canada)	1	2	2	1		1		1	1			1	10
New South Wales (Australia)	1	1	1	1	1	1	1	1			1	1	10
Massachusetts (USA)	1	1	1	1	1			1	1	1	1		9
Ontario (Canada)	1	1	1	1	1	1		1			1	1	9
China	1		2	2	1				2			1	9
USA (Federal)		1	1	1				1	1			1	9
Japan	1	1	1	1		1							5
Russia	1	1	1						1				4
Total	38	38	38	37	32	27	19	29	32	21	26	25	

Researchers looked at the frequency of mentions in curriculum documents across 5 categories - Competency inclusion, Competency identification, Competency progressions, Competency pedagogies and Competency assessments. A "3" (blue) indicates that the jurisdiction has identified that competency in 3 categories, a "2" (orange) indicates that the competency was identified in two categories and so on.

Differing approaches to teaching and learning

In terms of the kinds of pedagogy/ instruction needed today to develop both the foundational literacies and the kinds of desirable dispositions listed in Table 1, the educational world has become unhelpfully polarised.

One group, broadly those who might see themselves as traditional, tends to argue for the teacher's role in transmitting knowledge and to favour didactic methods. The other, typically seen as progressive, argues for student-led approaches, such as problem-based learning. With Guy Claxton (Claxton and Lucas, 2015) I have suggested

that there is a middle way; that these are false binary positions.

In such a mid-position 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?

As argued so far, there is emerging agreement as to the answers to the first two questions, with continuing discussions about the other six.

There are many other questions we could pose with regard to pedagogy or instruction, and many sources of evidence on which educators can draw (Hattie, 2008; Committee on Developments in the Science of Learning, 2000; Coe et al, 2020). Importantly, any teacher reaching for such guidance will need to consider not just which teaching methods promote success in terms of typical examinations, but

which methods **also** cultivate the kinds of dispositions or capabilities young people need. Ideally, methods that effectively promote both outcomes will be chosen.

Skills are what matter in life. Skills are the 'connective tissue' between knowledge and dispositions.

A misunderstanding of the role of skills in learning

There is much nonsense talked about skills today.

By those who see the acquisition of knowledge as the main purpose of education, an emphasis on skills is often portrayed as an attempt to dumb down or distract schools from their core purpose.

By those who see dispositions and capabilities as being centrally important, there is a temptation to hold fast to bigger concepts, such as creativity or collaboration, without recognising that they are in reality made up of aspects of knowledge and clusters of skills.

Knowledge and dispositions are not polar opposites, just different ways of categorising what we can learn. The 'currency' of both is skills. Skills are what matter in life. Skills are the 'connective tissue' between knowledge and dispositions. As we practise a skill in different contexts we become more competent, confident and capable, until it becomes a disposition, something we are disposed to do. Some examples, from simple to more complex, might include

- planning an essay;
- delivering a speech;
- critiquing an argument;



trying different approaches when faced with a tricky problem;

using your common sense when your satnav takes you to a cul-de-sac not yet updated in its software;

reading the mood of those with whom you are working;

facilitating a workshop where you are a content expert; or

facilitating a workshop when you have only a basic knowledge of the context but can transfer facilitation skills learned in other contexts to the task at hand.

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

And, yes, recalling decontextualised information in a pencil and paper examination is a skill, but not one that adults need to use much in a digital age.

The deeper your knowledge and the more you practise your skills in a variety of contexts, the more capable you become. Dispositions are clusters of skills which have been practised so well that they have become habitual; you are routinely disposed to deploy them; and skills are the mechanism by which knowledge is applied and dispositions are lived out.

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

So, to assessment. 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. This is especially true in upper secondary schools as students reach the age when they move on to university, vocational training or employment. 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.



Most education systems are seriously in need of attention if they are to be fit for purpose.

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.

Most education systems are seriously in need of attention if they are to be fit for purpose. Curricula are changing, debates are at least being had about pedagogy (Griffin, McGaw and Care, 2012; Vincent-Lancrin, et al, 2019) but, despite some promising initiatives, assessment needs some serious rethinking.

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

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

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

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.

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.

Traditional areas, literacy, maths and science continue to require considerable content to be tested, while newer areas such as citizenship, engineering, 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.⁴

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

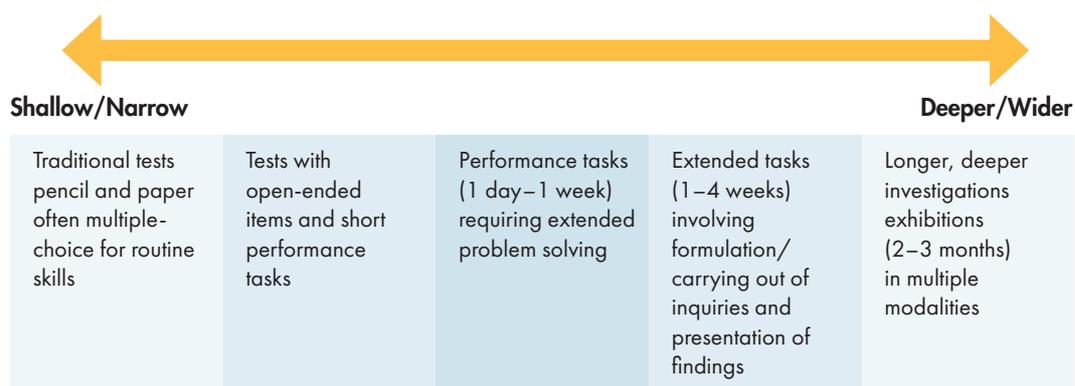
Most tests used in schools still rely on paper and pencil. They examine aspects of knowledge and routine skills. They test students' ability to **remember** and **write** about something, rather than **apply** or **do** the thing they have been learning. Concepts and skills are tested in individual subjects 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), or the external moderation by an accountability body that keeps levels of achievement very similar year on year (as with GCSE in England). 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’ (High Resolves, 2020, p 16) 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 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 or, frequently ineffectively, of employers. 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’ (Wiliam, 2007, p 248) 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’ (Hipkins and Cameron, 2018, p 22).

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);

increasing irrelevance to employers (Harvard Business Review, 2015);⁵

distracting from the huge importance of assessment for learning and assessment as learning (Birenbaum et al, 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.

(p 39)

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.

(p 11)

Dubious validity for many users

Assessment serves many purposes, including the following.

It certifies, selects and credentials students for universities and colleges.

It is a sifting mechanism for employers.

It gives teachers information on the progress of their students.

It gives students actionable feedback on their progress and suggests potential next steps.

Across the world, however, there is a crisis of validity, with growing dissatisfaction from each of the main users.

Universities and colleges

Universities and colleges find the grades or scores they are provided with too crude to be helpful, so that many are creating consortia to work with schools to provide more rounded information. The Mastery Transcript Consortium,⁶ the New York Performance Standards Consortium⁷ and the Comprehensive Learner Record,⁸ in the USA, and New Metrics for Success, in Australia,⁹ are indicators of a growing unease with the status quo.

Employers

Employers are frustrated that the current crop of academic and vocational qualifications leave them under-informed about potential employees (Education Council, 2020; Confederation of Business Industry (CBI), 2019). Many employers are now qualification-blind in their recruitment. In England, Rethinking Assessment has identified many examples of, predominantly, larger organisations that operate in this way, including Apple, Bank of America, BBC, the Civil Service, Clifford Chance, Google, The Guardian, Hilton, Microsoft, Penguin Random House, PricewaterhouseCoopers (PWC) and Starbucks.

Many employers now develop their own approaches to assessing potential employees. Often these are 'strength-based' aptitude tests, looking to see what capabilities and values candidates have to better enable them to work productively with others, seeking to establish a more balanced scorecard than mere exam grades.

Teachers ... are concerned variously about the way that tests privilege certain subjects over others, especially 'academic' over practical, and how an emphasis on memorisation can lead to shallower and less enjoyable learning, especially at upper secondary level

As Professor Tristram Hooley, Chief Research Officer of the Institute of Student Employers in England, puts it,

*Most employers don't worry if a candidate knows a little bit less about theories of population migration or the nineteenth century novel. But they will care a lot about candidates' ability to learn, to think on their feet, to be resilient in the face of knock backs, and so on.*¹⁰

The old narrative of working hard, getting good grades at school, going to a good university and securing a well-paid job is increasingly fractured. Employers are becoming aware that, 'when it comes to predicting job performance, aptitude tests are twice as predictive as job interviews, three times as predictive as job experience, and four times as predictive as education level'.¹¹

Teachers

Teachers have had rising degrees of dissatisfaction with the status quo since the millennium. They are concerned variously about the way that tests privilege certain subjects over others, especially 'academic' over practical, and how an emphasis on memorisation can lead to shallower and less enjoyable learning, especially at upper secondary level. This was evident in England two decades ago.

When passing tests is high stakes, teachers adopt a teaching style which emphasises transmission teaching of knowledge, thereby favouring those students who prefer to learn in this way and disadvantaging and lowering the self-esteem of those who prefer more active and creative learning experiences.

(Harlen and Deakin Crick, 2002, p 4)

Wherever you are in the world, the COVID-19 pandemic has provided a dramatic interruption of normal assessment activity. PISA's 2021 tests are currently rescheduled until 2022. Across the world, school examinations for 18-year-olds or 19-year-olds have been cancelled, postponed or simplified.¹² In many cases these changes have required students to rely on teacher-assessed grades. While this can be seen as a positive development (inviting innovation in methods), in practice it has caused additional stress among teachers who may not yet be assessment literate enough to undertake such testing without an appropriate infrastructure of moderation and training, along with equitable appeals processes.

Students

Students are increasingly unsettled. In one part of their world they have moved from an era of television programs to be watched at set times, to unlimited on-demand consumption of You-Tube, TikTok and streaming services; from books which needed to be learned, to an Internet which can be searched. Not so their examinations, which mostly require pencil and paper completion on a set date and considerable feats of memory.

When it comes to high-stakes assessment, there is widespread and ongoing stress among students, as this blog¹³ on the website of Ofqual (The Office of Qualifications

and Examinations Regulation) in England highlights. In *The Testing Charade*, Daniel Koretz quotes an alarming letter from New York principals to parents.

We know that many children cried during or after testing, and others vomited or lost control of their bowels or bladders. Others simply gave up. One teacher reported that a student kept banging his head on the desk ...

(Koretz, 2017, p 2)

Educational jurisdictions

An educational jurisdiction's performance is also judged through international assessments. Assessments are used as a means by which society rates, often in very limited ways, the performance of its schools. Using tests such as the Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS), the success of individual jurisdictions can be compared internationally. These have a powerful impact on both what is tested and how it is evidenced, but that is beyond the scope of this discussion.

Can dispositions be measured?

In the last few decades we have made real progress in understanding how best to evidence dispositions more generally (Soland et al, 2013; Darling-Hammond, 2017; Siarova et al, 2017; Care et al, 2018). In some cases real progress is being made in developing useful standard measures of specific aspects of some key dispositions, for example of 'grit' (Duckworth and Quinn, 2009).

Interestingly, it is through tests like PISA that we have been able to make significant breakthroughs in our understanding of how two key dispositions/competencies, collaborative problem solving¹⁴ and creative thinking¹⁵, can be assessed in an online test. (I have been involved in helping to shape the second of these two tests.)

We have been assisted in this process by advances in assessment technology. For example, evidence-centred design, a way of creating assessments that better demonstrate how test-takers' inferences are made and their reasoning is developed as they approach assessment tasks, is a promising approach.



Table 4. Methods of evidencing progress in creativity (Lucas and Spencer, 2017, p 160)

Student	Teacher	Real-world	Online
Real-time feedback	Criterion-referenced grading	Expert reviews	Reliable, validated online tests
Photos	Rating of products and processes	Gallery critique	Digital badges
Self-report questionnaires	Structured interviews	Authentic tests eg displays, presentations, interviews, podcasts, films	E-portfolios
Logs/diaries/journals	Performance tasks	Exhibitions	
Peer review	Capstone projects		
Group critique			
Badges			
Portfolios			

In my own work in the UK and in Australia, working with schools and school systems and drawing on a wider OECD study (Vincent-Lancrin, et al, 2019) with which I was involved exploring the assessment of creativity, I have found that a clear understanding of what creativity is, along with an understanding of learning progression, is a necessary starting point. Then, provided a range of different perspectives are acknowledged, it is possible to provide students and teachers with robust evidence of progress over time (see Table 4).

Importantly, we need multimodal assessment to gain an accurate picture, using perspectives from at least three columns in Table 4.

However, we have a way to go yet. As Daniel Willingham reminded us in 2013, in his blog,

*we're far from agreed-upon measures. Just how big a problem is that? It depends on what you want to do. If you want to do science, it's not a problem at all. It's the normal situation.*¹⁶

In 2016 the journal *Applied Measurement in Education* compiled a special issue focusing on the assessment of so-called 21st century skills.¹⁷ It focused on four types of dispositions: collaborative problem solving; complex problem solving; digital and information literacy; and creativity, to which I contributed our research at the University of Winchester, (Lucas, 2016). In the spirit of scientific enquiry, the issue focused on both what we do know and what we do not yet fully understand. It offered some promising approaches, some of which are already being used by PISA.

Just as these days few contest the notion of the learning sciences as a valid lens to explore teaching, so we need a similar shift in building the science of assessment. I'll say more about this in the final section.

Revisiting the purposes of assessment

Assessments must fully represent the competencies that the increasingly complex and changing world demands. The best assessments can accelerate the acquisition of these competencies if they guide the actions of teachers and enable students to gauge their progress.

(Gordon Commission, 2013, p 7)

Over the last few decades we have progressively lost our way with educational assessment. What we assess grows ever further away from what we want young people to be able to know, do, be and become in the complex world in which they live today. The focus of most systems is on summarising rather than understanding, recalling rather than applying, noticing deficiencies rather than celebrating strengths. We seem happier using numbers rather than narratives, keener on judging rather than prompting improvement.

We seem happier using numbers rather than narratives, keener on judging rather than prompting improvement.

Previous attempts to rethink assessment

The Assessment Reform Group in the UK made ground-breaking progress in its exploration of the value of assessment for learning between 1996 and 2010, something which has subsequently spread across the world, albeit always in tension with the tendency of systems to prefer summative data, (Birenbaum et al, 2015). For a long while (Black and Wiliam, 1998) we have known that formative assessment is effective in promoting improvements in student learning.

Twenty years ago the Committee on the Foundations of Assessment in the USA (National Research Council, 2001) considered the degree to which advances in the cognitive sciences were impacting on educational assessment. The central problem it identified is that ‘most widely used assessments of academic achievement are based on highly restrictive beliefs about learning and competence not fully in keeping with current knowledge about human cognition and learning’ (p 1). The report explores many of the then-current kinds of assessment and exposes these to rigorous scrutiny. It concludes with a vision of assessment that has still not been realised in education:

In the future envisioned by the committee, educational assessments will be viewed as a facilitator of high levels of student achievement. They will help students learn and succeed in school by making as clear as possible to them, their teachers, and other education stakeholders the nature of their accomplishments and the progress of their learning.

(p 292)

In 2006 a group of states in Canada worked with Lorna Earl and Steven Katz to, as they described it, ‘rethink classroom assessment

with purpose in mind'. The document provides a framework for thinking about the **purposes** of assessment. Its reminders about the distinctions between assessment **for**, **as** and **of** learning are clear and succinct:

Assessment for learning is designed to give teachers information to modify and differentiate teaching and learning activities. It acknowledges that individual students learn in idiosyncratic ways, but it also recognizes that there are predictable patterns and pathways that many students follow.

Assessment as learning is a process of developing and supporting metacognition for students. Assessment as learning focusses on the role of the student as the critical connector between assessment and learning.

Assessment of learning is summative in nature and is used to confirm what students know and can do, to demonstrate whether they have achieved the curriculum outcomes, and, occasionally, to show how they are placed in relation to others.

(Earl and Katz, 2006, p 13–14)

The Gordon Commission in the USA in 2013 made a number of key recommendations, about designing and implementing assessment that supports a more ambitious and expansive vision of education. It is vitally important, the Commission argued, that assessments

best represent the kind of learning students will need to thrive in the world that awaits them beyond graduation (p 8) ... Assessments must advance competencies that are matched to the era in which we live. Contemporary students must be able to evaluate the validity and relevance of disparate pieces of information and draw conclusions from them.

(p 9)

In Australia, at the same time as the Gordon Commission, the Australian Council for Educational Research (ACER) undertook a review of educational assessment, (Masters, 2013). Geoff Masters reminds us that 'the fundamental purpose of assessment is to establish where learners are in their learning at the time of assessment' (p 5–6). In a thoughtful and prescient overview Masters points to the folly of age-related testing, the failure of assessment to enable effective differentiation in teaching, the way in which it supports 'traditional approaches to schooling, including the assembly-line model, whole-class teaching, age-based curricula' (Masters, 2013, p 4), its ignoring of important life skills such as collaborative working, and its pedestrian use of technology. Not much has changed in the eight years following the ACER review.

The COVID-19 pandemic, as well as requiring educational jurisdictions to rethink their approach to assessments, at least temporarily, has forced educators to think more carefully about issues of equity as we recover from the pandemic.

Assessments, if used properly, can help us to mitigate the impacts of the COVID-19 pandemic for years to come. If used improperly, assessments may waste precious instructional time and resources, worsen inequities, reinforce misperceptions as to sources of inequity, and impede sound education policy.

(National Academy of Education, 2021, p 13).

New Metrics for Success¹⁸ at Melbourne University in Australia, the Brookings Institution in the USA¹⁹ and Rethinking Assessment²⁰ in England are three examples of organisations trying to find saner solutions to many of the issues noted in these earlier attempts to reimagine assessment. Also we are starting to move with an urgency and a focus on practical solutions.



The science of assessment

Over the last two decades there has been a step change in the way in which evidence about learning is used by schools and within school systems; the science of learning is a widely accepted concept in education.

However, despite some occasional thoughtful publications (National Research Council, 2001), there has not been the same culture shift among teachers and policy-makers with regard to the science of assessment. A Google search on ‘science of learning’ produces 1,840,000,000 results, while one on ‘science of educational assessment’ yields 655,000,000, the majority of which on closer scrutiny turn out to be about the assessment of science education.

In a recent paper exploring implications for educational practice of the science of learning (Darling-Hammond et al, 2020) it is noteworthy that assessment does not merit a discrete section (despite the fact that the paper is written by a team including an assessment expert). Notwithstanding this, there are clear messages to be heeded.

Many schools that have been particularly successful in reducing opportunity and achievement gaps for traditionally marginalized students – producing high graduation and college success rates – have adopted mastery-oriented performance-based assessments that build higher order thinking and performance skills, collaboration and communication skills, motivation and engagement, and a host of co-cognitive skills such as self-regulation, executive function, resilience, perseverance and growth mindset (p 121) ... Performance assessments that encourage higher order thinking, evaluation, reasoning, and deep understanding are themselves tools for learning.

(p 122)

Along with an understanding of the purpose of any activity to evidence performance and progress in education, these kinds of observations form the bedrock of the assessment sciences that will help us take better decisions. To return to the metaphor of fishing with which I began this paper, the science of assessment will help us to cast the right kinds of nets.

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’ (See page 10).

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.

A focus on collaborative rather than just individual performance.

Authenticity

Increasing interest in strengths-based approaches, especially from employers.

The need to design better performance-based assessments.

A move towards scenario-based, authentic assessment.

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.

The desirability of international benchmarking.

Three things emerged at the metalevel. Systems need to decide what they value before they determine what they seek to evidence. An increasingly scientific approach to the field of educational assessment is required. While the direction of travel is increasingly clear, progress towards it is glacially slow.

Promising practices from across the world

*Mā te kimi ka kite, Mā te kite ka mōhio, Mā te mōhio ka mārāma.
Seek and discover. Discover and know. Know and become enlightened.*

(Maori saying)

There are many examples from which we can learn as we rethink the purposes and practices of assessment. As the Education Council (2020) reminded us in its review of senior secondary pathways into work in Australia, we need to learn from those who are actually trying things out, ‘Demonstration projects need to have greater influence on the traditional core of how we measure educational success’.

Interrogating practices

At Rethinking Assessment (RA) in England, we have been exploring a number of questions to better understand the

nature of the problems with which we are grappling, as we explore promising international demonstrations of what might be adopted in England. We have developed two sets of questions, one to do with knowledge and skills (see Table 5), and another relating to dispositions and skills (see Table 6).

Skills, as I have argued earlier, are the connective tissue between knowledge and dispositions.

An overarching question here concerns just how much knowledge we think students need to acquire at school and what kind of knowledge that is. Across the world, and reinforced by PISA’s focus, literacy, numeracy and science are generally considered to be foundational.

If the answer to this question is ‘less than we do now’, as many of us believe to be the case, then we will need to understand any possible consequences for the curriculum in schools of reducing the core focus of assessment.

At a more nuanced level we might want to look at the science curriculum to see which concepts are more relevant than others given the size of the field; with maths we might wish to re-emphasise content, prioritising, for example, statistics over some aspects of trigonometry. We might want to weigh the benefits of interdisciplinary knowledge, either expressed as, for example, STEM (Science, Technology, Engineering and Maths) or STEAM (Science, Technology, Engineering, Arts and Maths), or through an organisation of the curriculum into projects requiring more than one discipline. Both of these approaches are increasingly part of university life²¹ but surprisingly absent from schools.

An overarching question here concerns just how much knowledge we think students need to acquire at school and what kind of knowledge that is.

Table 5. Some guiding questions for RA when evidencing knowledge and skills

Knowledge and skills	<ol style="list-style-type: none">1. What is the core knowledge and skills that students should learn?2. How much of a student's knowledge and skills needs to be evidenced?3. What consequences will continue to follow from not assessing some aspects of knowledge and skills?4. How best can we evidence understanding?5. How best can we evidence the application of knowledge and skills in familiar situations?6. How best can we evidence the application of knowledge and skills to new situations?7. How best can we evidence interdisciplinary knowledge?8. How best can we evidence practical knowledge and skills?9. How best can we capture progress in the acquisition and application of knowledge?10. Who can we learn from to help better evidence the development and application of the knowledge students need today?
-----------------------------	---

Table 6. Some guiding questions for RA when evidencing dispositions and skills

Dispositions and skills	<ol style="list-style-type: none">1. In addition to knowledge, which dispositions should be cultivated in students?2. Which dispositions are both learnable and useful to students?3. What are the benefits of assessing dispositions?4. How technically reliably can different dispositions be evidenced?5. How practical are assessments of dispositions to deliver?6. What will the unintended consequences be of assessing certain dispositions?7. What consequences will continue to follow from not assessing dispositions?8. Who can we learn from to help better evidence the cultivation and demonstration of the dispositions students need today?
--------------------------------	---

In terms of potential assessment methods, Howard Gardner's words are still powerful.

Why talk about performances of understanding? So long as we examine individuals only on problems to which they have already been exposed, we simply cannot ascertain whether they have truly understood. They might have understood, but it is just as likely that they are simply relying on a good memory. The only reliable way to determine whether understanding has truly been achieved is to pose a new question or puzzle – one on which individuals could not have been coached – and to see how they fare.

(Gardner, 2006, p 34)

With dispositions and skills the questions necessarily have a different focus (see Table 6).

In terms of potential candidates for assessment, PISA offers us collaborative problem solving and creative thinking as two concepts that it has determined to be sufficiently robust and capable of being assessed reliably and validly, albeit on a relatively short computer test complemented by self-reported data.

Collaborative problem solving is only one aspect of collaboration. We do not just learn together to solve problems; we collaborate to generate new ideas, to make art, or to improve the contributions of others through well-chosen feedback and so forth.

The disposition of creative thinking as defined by PISA (OECD, 2019) essentially includes two clusters of skills, which typically are referred to as ‘creativity’ and ‘critical thinking’, both divergent and convergent thinking. We might wish to see both of these better valued and evidenced.

The Australian Council for Educational Research (Scouler et al, 2020) has made encouraging progress, in defining and then developing methods for evidencing creativity (creative thinking and critical thinking) and collaboration, when developing capabilities within the Australian Curriculum.

Many curriculum frameworks also increasingly stress Communication as a candidate for assessment. If this is seen as interdisciplinary, ubiquitous almost, then there is an argument for including it under the broad heading of dispositions (although it could equally be seen as a foundational literacy). Educators in the UK have been focusing on the development and assessment of oracy in its widest sense and have developed a set of benchmarks to facilitate more precise understanding of student progression, (Voice 21, 2019).

We might also want to look at grit, given the work that has gone into developing ways of assessing it.

In terms of evidencing dispositions, Rosemary Hipkins reminds us that, ‘Only when students are offered rich opportunities to demonstrate their capabilities will we know what they are actually capable of’ (Hipkins, 2018, p 22).

Sandra Milligan and colleagues (2020a, p 18) have elegantly captured the complexity of these kinds of assessment processes as a poem (see Box 1).

New kinds of nets

To return to the fishing metaphor with which I began this paper, we need better nets to identify the full range of young people’s talents.

The following is a selection of promising examples, loosely organised into categories.

Psychometric tests

In the main these are self-reported online tests or apps often used to evidence an aspect of character, wellbeing or metacognition.

Angela Duckworth’s grit scale²²

The Values in Action (VIA) Survey of Character Strength²³

The Harvard Human Flourishing app²⁴

The DESSA social and emotional learning assessment²⁵

Carol Dweck’s Growth Mindset assessment²⁶

The Metacognition Awareness Inventory²⁷

These tests as yet have varying degrees of reliability and validity.

Box 1.

Poem: Assessment of complex capabilities

Assessment is a process
of systematically observing what people say, do, make or write
during a relevant performance
which requires proficiency in the competence of interest
and using these observations as evidence
to support an overall judgment
about the position of the person
on a standard scale of expertise from less expert to more
indicating what they know and can do
and what they need to learn next
with a sufficient degree of precision
to allow recognition and reporting of the level of attainment.



KIPP schools²⁸ (Knowledge is Power Program) in the USA, for example, have adopted approaches to developing and assessing character, using resources such as the playbooks provided by CharacterLab.²⁹

Smart multiple choice

While some multiple choice tests can be reductionist and focus on recall or simple computations, others can be a way of evidencing aspects of dispositions such as critical thinking:

Raven's Progressive Matrices Test³⁰

California Critical Thinking Skills Test³¹

The Partnership for Assessment of Readiness for College and Careers (PARCC) in the USA has been moving towards greater freedom for students to construct their own answers rather than rely on predetermined responses.

The Mission Skills Assessment³² is an interesting example of a multiple choice test looking to evidence a wider set of constructs – creativity, curiosity, ethics, resilience, teamwork and time management.

Performance-based assessment

Performance-based assessment is a broad field encompassing traditional approaches from the Viva to AI simulations. Other examples include tests of proficiency in the arts and sports. Increasingly it is being used to enable students to perform tasks or activities that are meaningful and engaging. Performance also includes exhibitions, presentations and debates, sometimes associated with pedagogies such as project-based and problem-based learning.

Alelo's Oral Language Simulation³³ is a recent example of a computer-based simulation designed to measure not only proficiency in a foreign language. The program allows a student to interact directly with an avatar in a variety of languages.

Both the PISA test of collaborative problem solving and of creative thinking are performance-based assessments.

The Victorian Curriculum and Assessment Authority's online tests of Critical and Creative Thinking are examples of scenario-based tests of performance.³⁴

A recent European Union overview of assessment practices (Siarova, Sternadel and Mašidlauskaitė, 2017) suggests that

Performance-based assessment has the potential to measure and foster wide-ranging competences and higher-order skills, since it encompasses different assessment techniques and integrates a feedback mechanism. The key strengths of performance-based assessment include its focus on the learners' personalised needs, clear definitions of the learning goals, and timely feedback.

(p 8)

Extended investigations

Extended investigations are increasingly seen by researchers (Soland et al, 2013; Conley and Darling-Hammond, 2013; Hipkins and Cameron, 2018) as the kinds of assessments needed to best evidence higher-order thinking skills present in many dispositions.

Many extended investigations are or contain elements of performance-based assessment.

In addition to a growing number of schools using such approaches across the world, there are externally validated examples to consider.

The Extended Project Qualification³⁵ (EPQ), in England and Wales, enables students to undertake an investigation in the context of a project topic they have selected. Outcomes can be a design, performance, report, dissertation or artefact. An EPQ ‘counts’ for half the points value of an A Level for university entrance.

The Extended Essay in the International Baccalaureate (IB) is an independent, self-directed piece of research, finishing with a 4,000-word paper. The Extended Essay is presented as practical preparation for undergraduate research – an opportunity for students to investigate a topic of personal interest to them taking an interdisciplinary approach. Although encouraging extended investigation, these essays are writing about things rather than demonstrating capabilities or dispositions.

The South Australian Certificate of Education extended research projects³⁶ can be used to explore aspects of the Australian general capabilities. This assessment is currently in an interesting phase of further development.³⁷

The New South Wales Personal Interest Projects,³⁸ as the name suggests, offer the potential for extended investigation

but, like many such investigations, are assessable only in a written essay format.

Alongside the four methods illustrated above, there are two other dimensions worth noting, to do with the availability and timeliness of assessments, micro-credentialling and on-demand tests.

Micro-credentialling

Micro-credentials, sometimes referred to as digital badges, take an idea long used by scouts, guides and other informal youth organisations and bring it into the digital age. Aspects of a larger concept, such as a disposition, are reduced to a small number of skills and ‘badged’ up to enable students to acquire credential in bite sizes.

Badges have the advantage, too, of providing a visual image of a student’s progress at a glance. The process of using digital badges in Franklin School in the USA is well described by David Niguidula (2020).³⁹ Badges are, in the jargon of assessment, ‘stackable’. That is to say they can be combined together to evidence many different assessment outcomes.

Digital Promise,⁴⁰ in the USA, has been focusing on developing a system of micro-credentials for use in the professional learning of teachers. Digital Promise succinctly articulates the benefits of micro-credentialling, including being

- bite-sized, focused on a specific, observable competency;
- subject-adaptable, adaptable to multiple subject areas;
- research-based, grounded in educational research;
- personal and timely – supporting professional growth;
- portable, can be shared as digital badges in online platforms;
- transparent, supported by publicly available, accessible content, including criteria for assessment;

performance-based, demonstrated through plans, work samples, reflections, observations, videos and peer and self evaluations.

In Europe there is an attempt to build micro-credentials into a reliable system of credit transfer between schools and universities (Futures, Andersen and Larsen, 2020).

In England there are already mechanisms for awarding credit for discrete achievements,⁴¹ often in alternative settings, allowing learners the opportunity to have their achievements formally recognised with a certificate each time a short unit of learning is successfully completed.

On-demand and online

In our learning lives outside school the idea of only being able to take a driving test, for example, on a set date would be laughed at. In most countries such tests require some theory (typically an online multiple choice test) and a practical demonstration on the road. The on-the-road test is, unsurprisingly, attempted when we and our driving instructor think that we are capable. In similar vein, only being able to use our memory rather than our ability to search and apply knowledge from the Internet or from notes or materials we have made while studying is increasingly perverse.

On-demand testing is already widely used in music, for example. While still at school many students take grade exams to assess the quality of their playing classical, jazz or rock music when they have reached a standard that they and their teachers deem them to be ready for.

On-demand testing more generally in education makes practical sense and a number of educational jurisdictions are experimenting with it, for example the Victorian Curriculum and Assessment Authority (VCAA) in Australia.⁴²

Sugata Mitra suggests that, in future, assessment should precede teaching and learning should actively encourage critical thinking and consensus-building, using the Internet. He imagines a post-Pandemic classroom as follows.

Sessions usually start with a set of questions. In the pre-pandemic times, this would have been called a test. Tests were usually given after the 'teaching and learning' were over. Not so anymore. Sessions can start with tests. The children have no idea what the answers might be, they haven't been 'taught'. But they can look up things on the Internet and talk to each other. When the answers come in, the teacher begins a discussion. She encourages the children to talk about their answers, sometimes, very occasionally, she adds a bit. They arrive at a consensus by the end of the session about what the answers are and why.

(Mitra, 2020, p 287)

Mitra's open style of teaching has an equivalent in assessment, open-book tests. Here students are able to have key texts with them as they answer a question. Open-book tests do not just test a student's ability to recall information but require a more critical and analytical and applied approach to answering questions.

Partly as the result of Covid-19, schools and universities have gone one stage further and moved to set **online** open-book tests. Typically, students are required to complete an assessment within a defined time period, for example 24 hours, with questions released to students online at a specified time and answers required online by the end of the timed period. While there are some important equity issues that need to be considered here (broadband strength, computer availability and the availability of quiet spaces in students' homes), the essence of such approaches could be applied and used within the school day.

Far from dumbing down the quality of answers, such examinations potentially offer opportunities for deeper learning. require different kinds of questions to assess interpretation, evaluation and critical thinking rather than knowledge recall. Frequently they start from scenario-based or problem-based questions that require students to apply knowledge rather than summarise information. As a consequence they cannot be gamed by cutting and pasting from the Internet.

In a thoughtful recent blog, *The searching questions that will allow us to rethink assessment*,⁴³ Guy Claxton suggests that we need not spend time worrying about end-of-school exams but instead could leave assessment – he calls this process MOECs (Methods of Evidencing Capability) – to a college or university or to an employer rather than undertaking it at the exit point from school.

Games-based assessment

Games-based assessment is in its infancy in schools. Nevertheless there are examples from which we can learn.

Keenville⁴⁴ is a formative game-based assessment initiative for 1st and 2nd graders in the state of Georgia, a collaboration between the Georgia education department, the Georgia Center for Assessment at the University of Georgia and FableVision Studios.

Posterlet is a game created by Stanford's Graduate School of Education to measure students' choices in seeking feedback and revising work while at the same time learning graphic design principles (Cutumisu et al, 2015).

Critical and Creative Tests developed by VCAA in partnership with (initially) ACER and now NFER in England, use a range of engaging scenarios to test students' critical and creative thinking skills.

Comparative judgement

Comparative judgement uses the principle that people are better at making comparisons **between** pieces of work than at making **absolute** judgements about their quality. First identified as a useful principle of assessment nearly a century ago (Thurstone, 1927), the addition of digital technology enables a kind of crowd sourcing of teachers' judgements of particular usefulness when reviewing written work. Several approaches are being developed by No More Marking in the UK.⁴⁵

Profiling

Profiling students' wider achievements at school, using some kind of record of achievements, is not new.⁴⁶ In the 1990s, for example, the UK government encouraged all secondary students to develop a National Record of Achievement, a portfolio of documents showcasing a student's academic and non-academic achievements, including GCSE certificates, certificates from extracurricular activities, school reports and anything else considered relevant.

These Records of Achievement did not work because the non-standardised elements were often too long-winded to be assimilated by others; it was not clear who was warranting the data; and there was not buy-in from employers. They were ideas before their time and before digital technology was freely available. Fast forward to today and a number of individual schools, groups of schools and external providers are experimenting with profiles.

In the USA, Panorama Education has developed online dashboards or profiles to illustrate the range of a young person's social and emotional and learning skills.⁴⁷

For many years students taking the International Baccalaureate (IB) have graduated with an IB Profile as well as their diploma. The IB Learner Profile contains ten attributes (see Figure 4).

Figure 4. The ten attributes of the IB Learner Profile

IB learner profile

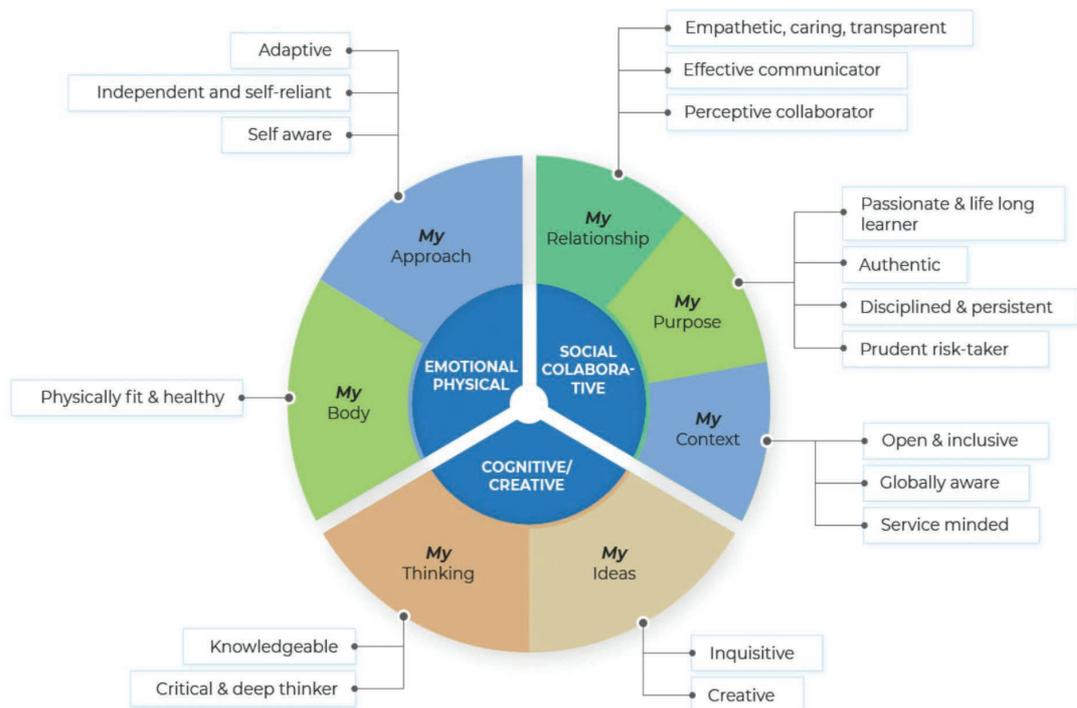
The aim of all IB programmes is to develop internationally minded people who, recognizing their common humanity and shared guardianship of the planet, help to create a better and more peaceful world.

As IB learners we strive to be:

<p>INQUIRERS We nurture our curiosity, developing skills for inquiry and research. We know how to learn independently and with others. We learn with enthusiasm and sustain our love of learning throughout life.</p> <p>KNOWLEDGEABLE We develop and use conceptual understanding, exploring knowledge across a range of disciplines. We engage with issues and ideas that have local and global significance.</p> <p>THINKERS We use critical and creative thinking skills to analyse and take responsible action on complex problems. We exercise initiative in making reasoned, ethical decisions.</p> <p>COMMUNICATORS We express ourselves confidently and creatively in more than one language and in many ways. We collaborate effectively, listening carefully to the perspectives of other individuals and groups.</p> <p>PRINCIPLED We act with integrity and honesty, with a strong sense of fairness and justice, and with respect for the dignity and rights of people everywhere. We take responsibility for our actions and their consequences.</p>	<p>OPEN-MINDED We critically appreciate our own cultures and personal histories, as well as the values and traditions of others. We seek and evaluate a range of points of view, and we are willing to grow from the experience.</p> <p>CARING We show empathy, compassion and respect. We have a commitment to service, and we act to make a positive difference in the lives of others and in the world around us.</p> <p>RISK-TAKERS We approach uncertainty with forethought and determination; we work independently and cooperatively to explore new ideas and innovative strategies. We are resourceful and resilient in the face of challenges and change.</p> <p>BALANCED We understand the importance of balancing different aspects of our lives—intellectual, physical, and emotional—to achieve well-being for ourselves and others. We recognize our interdependence with other people and with the world in which we live.</p> <p>REFLECTIVE We thoughtfully consider the world and our own ideas and experience. We work to understand our strengths and weaknesses in order to support our learning and personal development.</p>
--	---

Source: The Tacit Dimension: Understanding International Mindedness in Hong Kong International Baccalaureate Diploma Programme Schools

Figure 5. The Xperiential Learner Profile



In fact the IB Learner Profile is not a personalised description of the degree to which a learner possesses these attributes or dispositions. It is rather a statement of intent, of the kind of all-round individuals the IB values and which, it hopes, universities across the world will also like.

Numbers of schools are trying to develop profiles of student achievement that are closer to the spirit of the Records of Achievement three decades ago. Heritage Xperiential Learning School in Doncaster, England is an example of this. Modelled on High Tech High School in California, The Xperiential Learner Profile⁴⁸ mixes dispositions with values and health, (see Figure 5). Like the IB Profile, Xperiential's is a statement of educational intent. It is currently exploring ways of providing data to evidence each of its components.

For a number of years Rooty Hill High School in Sydney, a school with which I have been privileged to collaborate, has been developing a dashboard, of which this is an early version, for each of its students (see Figure 6).

The dashboard combines a portfolio – a space for students to upload evidence of their progress using the language of Australia's general capabilities – with a goal-setting function, as well as an opportunity for individuals to compare their progress to others' in the school.

Upskill Me⁴⁹ is an example of a growing number of platforms offering school students a way of tracking and curating their activities, receiving digital badges along the way by use of an app. With Upskill Me, young people can also connect with employers and go to events to find job and work experience opportunities. Upskill Me serves as a record of achievement.

Standing back from the different nets

These examples are illustrative only of an exciting, emerging field. Back to the fishing metaphor; it is tempting to get too close to the nets and be dazzled by all the activity. It would be easy to be caught up in the novelty of some of the methods and forget to be clear about the purpose of any assessment.

Figure 6. Rooty Hill High School's Learner Hub

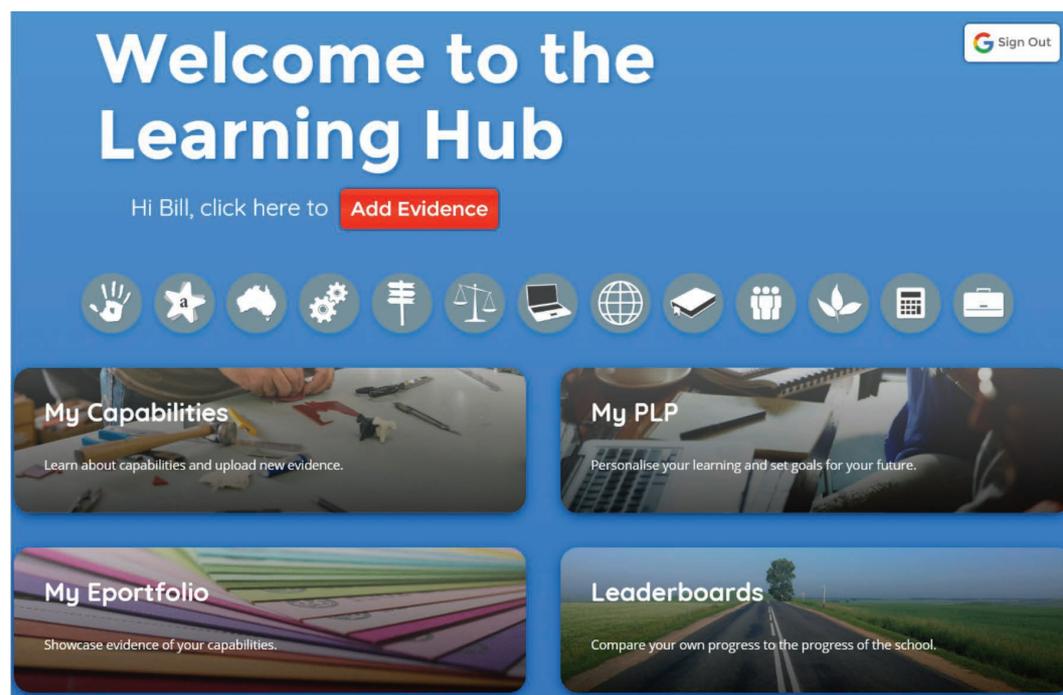


Table 7. Mapping possible assessment methods to a possible disposition, Creativity

Focus	Method	Reliability/Validity	Pros	Cons
Creativity	1. Victorian Curriculum and Assessment Authority (VCAA) online game-based	High, independently validated	Engaging for students. Useful feedback to teachers. Assessment as learning. Can be used summatively to show performance data over time for State.	Expensive to develop and validate. Need to quarantine some so that they can be used for benchmarking.
	2. Teacher observation of products and processes	Medium, depending on accuracy of scope and sequence and learning progression documents, moderation processes, professional support	More precise understanding by students of concepts. Enhanced teacher capability	

How best might we curate this array of approaches? Thus far I have loosely clustered them according to the **type of test**.

In Figure 3 we used the **continuum of shallow to deep learning** to map how more complex tests are needed to test deeper learning (Darling-Hammond, 2017).

Lorna Earl and Steven Katz (2006, p 17) suggest that we can view assessment processes according to their **purposes**, of

1. gathering information, eg, observations, tests, computer-based assessments, projects, rich tasks;
2. interpreting information, eg, developmental continua, rubrics, learning progressions, self-assessment, peer-assessment;
3. record-keeping, eg, observations over time, photographs or work, portfolios, digital badges; and
4. communicating, eg, demonstrations, presentations, exhibitions, records of achievement, profiles.

Alternatively, it might be easier to start from a list of potential dispositions (or aspects of knowledge, but these have been better analysed by others already) and cross map possible ways of evidencing these, along with their respective pros

and cons. Table 7 takes Creativity as an example of this process.

Table 7 includes just two examples, as being illustrative of the approach; there are many others.

In its recent project exploring the assessment of Australia’s general capabilities, the Australian Council for Educational Research (ACER) has developed a well-evidenced approach to their assessment using authentic, problem-based tasks. ACER’s assessment template

- measures multiple skills;
- is problem-based and authentic;
- is domain-orientated;
- maps to skill development levels.

(Scoular et al, 2020).

Whatever the specific ways of evidencing dispositions or capabilities, it is important to remember that research demonstrates ‘that there is no single method that would fully measure key competences and transversal skills, nor serve as a best practice for student assessment’ (Siarova, Sternadel and Mašidlauskaitė, 2017, p 8). As High Resolves (2020) has argued, we need to focus on better understanding the best combinations of multimodal assessments to select depending on context and desired outcomes.

Visible progress

The most important assessments that take place in a school building are seen by no-one. They take place inside the heads of students, all day long. Students assess what they do, say and produce, and decide what is good enough. These internal assessments govern how much they care, how hard they work, and how much they learn.

(Berger, Rugen and Woodfin, 2014, p 6)

A little over a decade ago, John Hattie published a ground-breaking book, *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*, (2008). It was remarkable in two ways. It dared to use the word ‘meta-analyses’, in a popular book for educators. More importantly, it lifted the lid on educational research for teachers across the world. In painstaking detail Hattie made the processes of learning visible, clear and actionable. For most people in education it was a game-changer. A similar paradigm shift is now needed in assessment.

We need to explore, in similarly nuanced and evidenced ways, the different ways in which we can use assessment to improve learning, and make these discussions visible to all, depending on the purpose we have in mind.

We need to make clearer the kinds of dispositions and capabilities that we want all young people to acquire, and make visible the processes by which we evidence student progress in acquiring them.

There are already educators beginning to do just this at a grass roots level, such as Ron Berger and colleagues, in Expeditionary Learning schools in the USA, whose words begin this section.

Building on earlier evidence

We are not starting from scratch. In Australia a Review undertaken by Geoff Masters in 2013 argued for a fundamental rethink about the purpose of assessments, that they should be seen as

having a single general purpose: to establish where learners are in their long-term progress, within a domain of learning at the time of assessment. The purpose is not so much to judge as to understand. This unifying principle, which has potential benefits for learners, teachers and other educational decision makers, can be applied to assessments at all levels of decision making, from classrooms to cabinet rooms.

(Masters, 2013, p 58)

The Gordon Commission, on assessment in the USA, spoke similarly.

In our vision of the future of assessment, the improvement of learning is its central purpose. It functions in dynamic interaction with curriculum and instruction, which themselves have the improvement of learning as its central purpose. Decisions about the form and content of assessment are informed by a socio-cultural perspective of learning,

curriculum and instruction and its results are used by both the teacher and the learner to guide future teaching and learning.

(Armour-Thomas and Gordon, 2013, p 19)

From an English perspective, Peter Hill and Michael Barber put it as follows.

Perhaps the most urgent need right now in the field of assessment is an overall conceptual framework and longer-term vision for its place and purpose within the triad of processes that lie at the heart of schooling.

(Hill and Barber, 2014, p 9)

The ‘triad of processes’ that underpin effective learning referred to are curriculum, pedagogy and assessment. In too many systems they have become disconnected.

A paradigm shift

The views cited above are just three of many but, as we saw in earlier sections, they represent a growing consensus about the direction in which we should be going. In this paper we have also both looked at what is wrong with educational assessment today and explored what could offer hope for developing better fishing nets in the sea of education.

The future is visual and digital

In my horizon-scanning of promising practices, it is increasingly clear that when it comes to evidencing the knowledge, skills and dispositions young people have acquired in their time at school, rather than a set of scores or grades, there is a move towards using Learner Profiles, Transcripts, Records of Achievement and Portfolios, alongside whatever standardised data is felt to be important. Figure 7 (Milligan et al, 2020a, p 24) is illustrative.

The Mastery Transcript Consortium (MTC)⁵⁰ in the USA, Figure 8, is another example of this trend.

The Figure 8 example is an early version and, in discussions with colleagues at MTC, I was interested to hear that, now that the transcript is gaining credibility with university admission staff, the number at the centre will be removed from future transcripts in favour of the more balanced transcript.

The field of profiling is evolving rapidly and there is as yet no single approach. The degree to which formal qualification scores or grades are prominent varies. Sometimes courses completed are included as an indication of a student’s dispositions or interests; in other examples there is an attempt to represent capabilities in some kind of spider or radar chart.

Figure 7. Prototype ARC Chrysanthemum Learner Profile

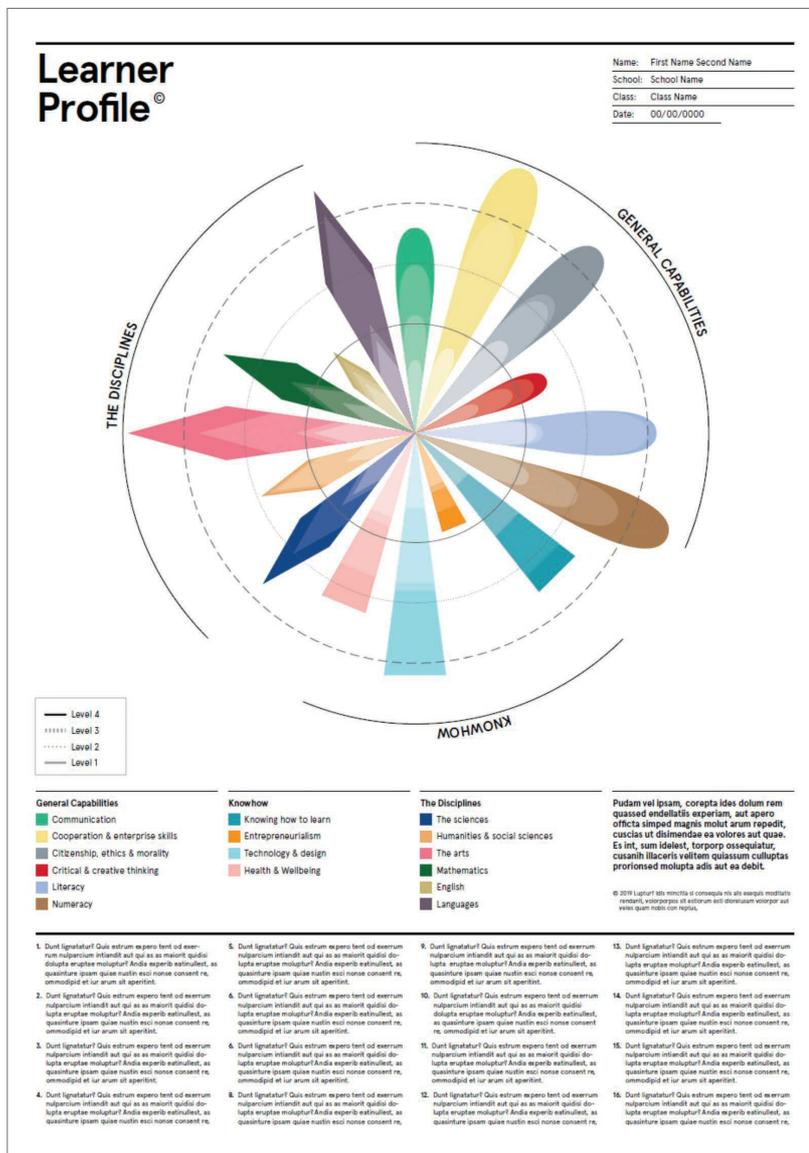


Figure 8. Mastery Transcript Consortium

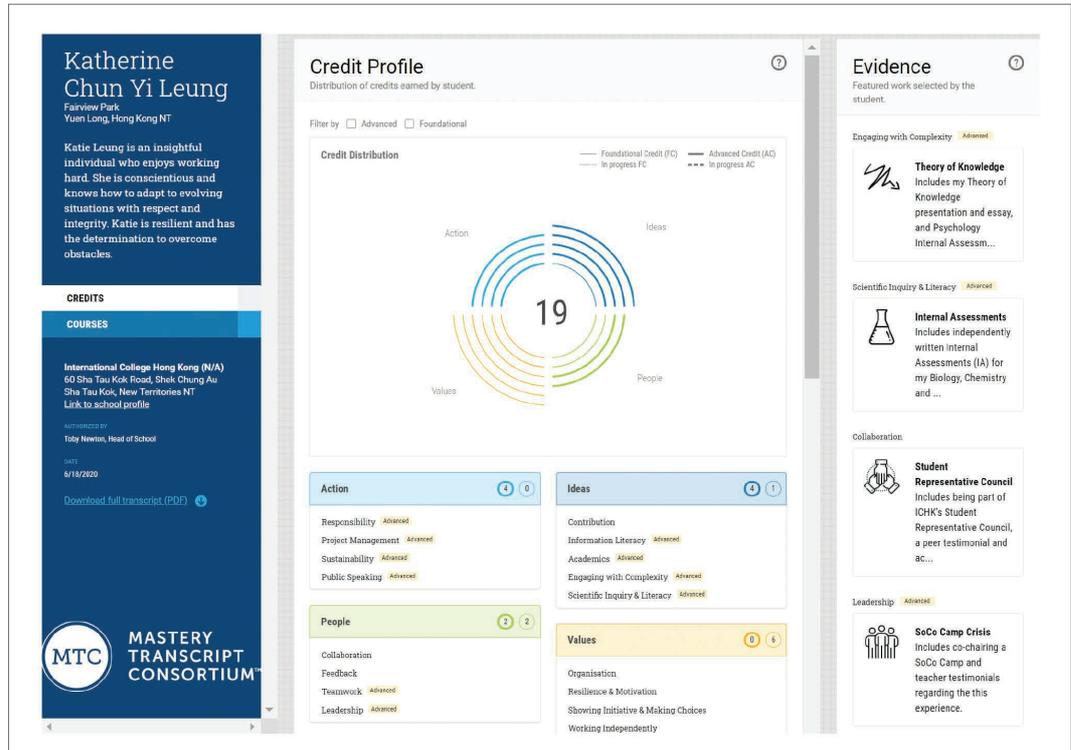


Figure 9. Big Picture Education Australia Profile

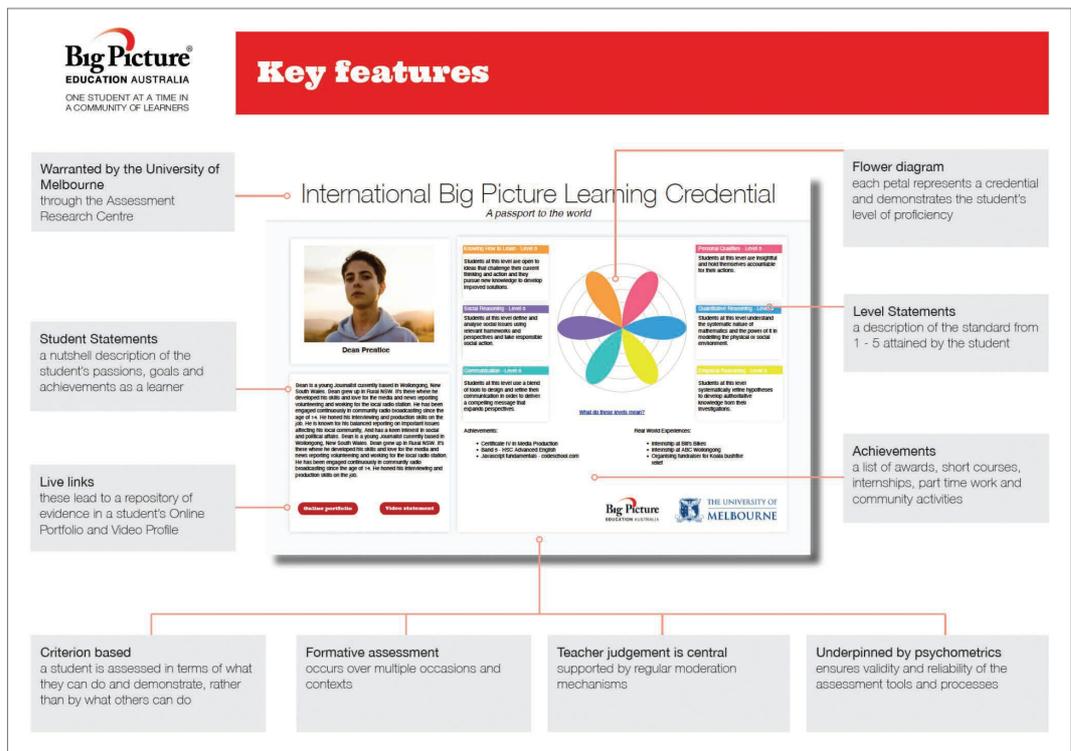


Figure 10. New directions in assessment



How student agency is exercised is also variable. Minimally there is a process of deciding what evidence should be foregrounded for a particular audience; in other cases a portfolio of best and most beautiful work is being curated over time by the learner.

A particular challenge for developers and consortia is ensuring that the approach is seen to be reliable enough. So, for example, the Big Picture example shown in Figure 9 is warranted by the University of Melbourne.

Time for a paradigm shift in educational assessment

We have a unique opportunity. The stakeholders are in agreement. The evidence is increasingly clear. Figure 10 brings together the direction of shift that the research has identified, and which innovative practitioners are prototyping.

New nets for a sea of opportunity

To return to the metaphor of fishing with which I began, it is time to value more kinds of fish, to know more about the fish we catch, and to use some new nets as we do so.

The evidence points to finding appropriate ways of valuing the skills that go to make

up the dispositions or capabilities of creativity, critical thinking collaboration and aspects of communication. At the same time we may wish to evidence less of what is currently assessed.

We need to have much more nuanced, strengths-based, multimodal descriptions of young people.

We need to use some of the many new methods being pioneered across the world, always seeking to make the processes of evidencing progress, in all aspects of learning, visible and evidence-based.

As we consider the inclusion of any new area, we will need to use evidence from the learning sciences to consider

its learnability;

its usefulness in life;

the validity, reliability and practicality with which it can be assessed; and

its likely positive impact on the development of more expansive curricula.

In addition, we need to get on with it **now** to reunite assessment with curriculum and pedagogy, from which it has become harmfully separated.

Endnotes

1. pz.harvard.edu/sites/default/files/AssessmentReimagined_Booklet_0.pdf
2. curriculumredesign.org/framework/
3. npdl.global/
4. issues.org/the-world-needs-students-with-interdisciplinary-education/
5. hbr.org/2015/07/ace-the-assessment
6. mastery.org/
7. performanceassessment.org/
8. imglobal.org/activity/comprehensive-learner-record
9. education.unimelb.edu.au/new-metrics-for-success
10. hepi.ac.uk/2019/08/14/why-employers-dont-care-about-qualifications/
11. criteriacorp.com/resources/definitive-guide-validity-of-preemployment-tests/what-are-pre-employment-tests
12. universityworldnews.com/post.php?story=2021011511142211
13. ofqual.blog.gov.uk/2019/03/08/what-is-it-like-to-experience-exam-stress-a-student-perspective/
14. oecd-ilibrary.org/docserver/9789264281820-8-en.pdf?expires=1615190774&id=id&accname=guest&checksum=022E0FA0656AA5AC791C134D42F5FB1B
15. oecd.org/pisa/publications/PISA-2021-creative-thinking-framework.pdf
16. danielwillingham.com/daniel-willingham-science-and-education-blog/measurement-of-non-cognitive-factors
17. *Applied Measurement in Education*, Volume 29, Issue 4 (2016) 21st Century Skill Assessment
18. education.unimelb.edu.au/new-metrics-for-success
19. brookings.edu/blog/education-plus-development/2017/08/31/new-data-on-the-breadth-of-skills-movement-over-150-countries-included/
20. rethinkingassessment.com/
21. For example nmit.ac.uk/ and <https://www.londoninterdisciplinarieschool.org/>
22. angeladuckworth.com/grit-scale/
23. viacharacter.org/
24. hfh.fas.harvard.edu/flourishing.app
25. apertureed.com/products-solutions/dessa-system-2/dessa-overview/
26. blog.mindsetworks.com/what-s-my-mindset
27. services.viu.ca/sites/default/files/metacognitive-awareness-inventory.pdf
28. kipp.org/approach/character/
29. characterlab.org/playbooks/
30. pearsonclinical.co.uk/Psychology/AdultCognitionNeuropsychologyandLanguage/AdultGeneralAbilities/Ravens-Progressive-Matrices/Ravens-Progressive-Matrices.aspx
31. insightassessment.com/article/california-critical-thinking-skills-test-family
32. bbk12e1-cdn.myschoolcdn.com/ftpimages/721/misc/misc_178206.pdf
33. alelo.com/
34. vcaa.vic.edu.au/assessment/f-10assessment/edstateap/Pages/cct-assessments.aspx
35. qips.ucas.com/qip/extended-project-qualification-epq#:~:text=The per cent20EPQ per cent20is per cent20a per cent20single, per cent20C per cent20report per cent20C per cent20dissertation per cent20or per cent20artefact.
36. sace.sa.edu.au/web/research-project
37. education.sa.gov.au/sites/default/files/sace-stage-2-review.pdf
38. educationstandards.nsw.edu.au/wps/portal/nesa/11-12/stage-6-learning-areas/hsie/society-culture/personal-interest-project
39. ascd.org/publications/educational-leadership/sept20/vol78/num01/The-Power-of-Digital-Badges.aspx
40. digitalpromise.org/initiative/educator-micro-credentials/
41. aqa.org.uk/programmes/unit-award-scheme
42. vcaa.vic.edu.au/assessment/f-10assessment/ondemandtesting/Pages/benefits.aspx
43. rethinkingassessment.com/rethinking-blogs/the-searching-questions-that-will-allow-us-to-rethink-assessment/
44. fablevisionstudios.com/keenville
45. nomoremarking.com/
46. As an early career teacher in England in the 1970s, I was seconded by the Oxford Certificate of Educational Achievement (OCEA), a partnership between the University of Oxford Examination Board and Oxfordshire Education Authority, to help develop a Record of Achievement that combined academic tests with evidence of extra-curricular and community learning and also included a large element of student self-assessment.
47. panoramaed.com/
48. heritagexperiential.org/about-us/learner-profile/
49. upskillme.io/
50. pilot-transcript.mastery.org/

References

- Armour-Thomas, E and Gordon, E (2013) *The Gordon Commission on the Future of Assessment in Education. Toward an Understanding of Assessment as a Dynamic Component of Pedagogy*, Educational Testing Service, Princeton, NJ.
- ASCL (2019) *The Forgotten Third: Final Report of the Commission of Inquiry*, Association of School and College Leaders, Leicester.
- Au, W (2016) 'Meritocracy 2.0: High-stakes, standardized testing as a racial project of neoliberal multiculturalism', *Educational Policy*, **30**, 1, p 39–62.
- Bennett, E (2013) *Preparing for the Future: What Educational Assessment Must Do*, Gordon Commission/Educational Testing Service, Princeton, NJ.
- Berger, R, Rugen, L and Woodfin, L (2014) *Leaders of Their Own Learning: Transforming Schools Through Student-Engaged Assessment*, Jossey-Bass, San Francisco, CA.
- Birenbaum, M, DeLuca, C, Earl, L, Heritage, M, Klenowski, V, Looney, A, Smith, K, Timperley, H, Volante, L and Wyatt-Smith, C (2015) 'International trends in the implementation of assessment for learning: Implications for policy and practice', *Policy Futures in Education*, **13**, p 117–140.
- Black, P and Wiliam, D (1998) 'Assessment and classroom learning', *Assessment in Education: Principles, Policy and Practice*, **5**, 1, p 7–73.
- Care, E, Anderson, K and Kim, H (2016) *Visualizing the Breadth of Skills Movement Across Education Systems*, Brookings Institution, Washington, DC.
- Care, E, Kim, H, Vista, A and Anderson, K (2018) *Education System Alignment for 21st Century Skills: Focus on Assessment*, Brookings Institution, Washington, DC.
- Care, E, Anderson, K and Kim, H (2019) *Assessment of Transversal Competencies: Current Tools in the Asian Region*, UNESCO, Paris.
- CBI (2019) *Getting Young People 'Work Ready': Our Vision of How Education Should Prepare Young People for the Modern World*, Confederation of Business Industry, London.
- Claxton, G and Lucas, B (2015) *Educating Ruby: What Our Children Really Need to Learn*, Crown House Publishing, Carmarthen.
- Coe, R, Rauch, C, Kime, S and Singleton, D (2020) *Great Teaching Toolkit: Evidence Review*, Evidence-based Education, Sunderland.
- Committee on Developments in the Science of Learning (2000) *How People Learn: Brain, Mind, Experience and School*, The National Academies, Washington, DC.
- Conley, D and Darling-Hammond, L (2013) *Creating Systems of Assessment for Deeper Learning*, Stanford Center for Opportunity Policy in Education, Stanford, CA.
- Cutumisu, M, Blair, K, Chin, D and Schwartz, D (2015) 'Posterlet: A game-based assessment of children's choices to seek feedback and to revise', *Journal of Learning Analytics*, **2**, 1, p 49–71.
- Darling-Hammond, L (2017) *Developing and Measuring Higher-Order Skills: Models for State Performance Assessment Systems*, Council of Chief State School Officers, Washington, DC.
- Darling-Hammond, L, Flook, L, Cook-Harvey, C, Barron, B and Osher, D (2020) 'Implications for educational practice of the science of learning and development', *Applied Developmental Science*, **24**, 2, p 97–140.
- Duckworth, A and Quinn, P (2009) 'Development and validation of the Short Grit Scale (GRIT-S)', *Journal of Personality Assessment*, **91**, p 166–174.
- Earl, L and Katz, S (2006) *Rethinking Classroom Assessment with Purpose in Mind: Assessment for Learning, Assessment as Learning, Assessment of Learning*, Manitoba Education, Citizenship and Youth, Manitoba.
- Education Council (2020) *Looking to the Future: The Report of the Review of Senior Secondary Pathways into Work, Further Education and Training*, Department of Education, Skills and Employment, Canberra.
- Education Policy Institute (2019) *'The Forgotten Third': A Rapid Review of the Evidence*, Association of School and College Leaders, Leicester.

- Elmore, R (2019) 'The future of learning and the future of assessment', *ECNU Review of Education*, 2, 3, p 328–341.
- European Parliament (2007) *Key Competences for Lifelong Learning: European Reference Framework*, Office for Official Publications of the European Communities, Luxembourg.
- Fullan, M (2021) *The Right Drivers for Whole System Success*, CSE Leading Education Series #1, February, Centre for Strategic Education, Melbourne.
- Futures, H, Andersen, T and Larsen, K (2020) *A European Approach to Micro-Credentials: Output of the Micro-Credentials Higher Education Consultation Group*, Publications Office of the European Union, Luxembourg.
- Gardner, H (2006) *Five Minds for the Future*, Harvard Business School Press, Boston, MA.
- Gordon Commission on the Future of Assessment in Education (2013) *To Assess, to Teach, to Learn. A Vision for the Future of Assessment*, Educational Testing Service, Princeton, NJ.
- Griffin, P, McGaw, B and Care, E (Eds) (2012) *Assessment and Teaching of 21st Century Skills*, Springer, New York.
- Gutman, L and Schoon, I (2013) *The Impact of Non-Cognitive Skills on Outcomes for Young People: Literature Review*, Institute of Education, University of London, London.
- Harlen, W (2005) 'Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes', *Research Papers in Education*, 20, 3, p 245–270.
- Harlen, W and Deakin Crick, R (2002) *A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation*, Evidence for Policy and Practice Information and Co-ordinating Centre, London.
- Hattie, J (2008) *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*, Routledge, Abingdon, Oxon.
- Heckman, J and Kautz, T (2013) *Fostering and Measuring Skills: Interventions that Improve Character and Cognition*, NBER Working Paper No 19656, National Bureau of Economic Research.
- High Resolves (2020) *The Future of Assessment: White Paper 2*, High Resolves, Sydney.
- Hill, P and Barber, M (2014) *Preparing for a Renaissance in Assessment*, Pearson, London.
- Hipkins, R and Cameron, M (2018) *Trends in Assessment: An Overview of Themes in the Literature*, New Zealand Council for Educational Research, Wellington.
- Howard, E (2020) *A Review of the Literature Concerning Anxiety for Educational Assessments*, Ofqual, Coventry.
- Koretz, D (2017) *The Testing Charade: Pretending to Make Schools Better*, University of Chicago Press, Chicago.
- Lai, R and Viering, M (2012) *Assessing 21st Century Skills: Integrating Research Findings*, Council on Measurement in Education, Vancouver, BC.
- Lamb, S, Maire, Q and Doecke, E (2017) *Key Skills for the 21st Century: An Evidence-Based Review*, Centre for International Research on Education Systems, Sydney.
- Lucas, B (2016) 'A five-dimensional model of creativity and its assessment in schools', *Applied Measurement in Education*, 29, 4, p 278–290.
- Lucas, B (2019) *Why We Need to Stop Talking About Twenty-first Century Skills*, CSE Seminar Series Paper 283, April, Centre for Strategic Education, Melbourne.
- Lucas, B and Spencer, E (2017) *Teaching Creative Thinking: Developing Learners who can Generate Ideas and Think Critically*, Crown House Publishing Ltd, Carmarthen.
- Lucas, B, Claxton, G and Spencer, E (2013) *Expansive Education: Teaching Learners for the Real World*, Australian Council for Educational Research, Melbourne.
- Masters, G (2013). *Reforming Educational Assessment: Imperatives, Principles and Challenges*. Australian Council for Educational Research, Melbourne. Also presentation for Open University, Maidenhead, UK.
- McEachen, J (2017) *Assessment for Deep Learning: A White Paper*, New Pedagogies for Deeper Learning, Ontario.

- Milligan, S, Luo, R, Hassion, E and Johnson, J (2020a) *Future-Proofing Students: What They Need to Know and How to Assess and Credential Them*, University of Melbourne, Melbourne.
- Milligan, S, Luo, R, Kamai, T, Rice, S and Keang, T (2020b) *Recognition of Learning Success for All: Ensuring Trust and Utility in a New Approach to Recognition of Learning in Senior Secondary Education in Australia*, Learning Creates Australia, Melbourne.
- Mitra, S (2020) 'Children and the internet: Learning in the times to come', *Journal of Learning for Development*, **73**, 3, p 286–305.
- National Academy of Education (2021) *Educational Assessments in the Covid-19 Era and Beyond*, National Academy of Education, Washington, DC.
- National Research Council (2001) *Knowing What Students Know: The Science and Design of Educational Assessment*, The National Academies Press, Washington, DC.
- Niguidula, D (2020) 'The power of digital badges', *Educational Leadership*, **78**, 1.
- O'Connell, M, Milligan, S and Bentley, T (2019) *Beyond ATAR: A Proposal for Change*, Koshland Innovation Fund. Accessed at apo.org.au/node/261456.
- OECD (2019) *PISA Creative Thinking Framework, Third Draft*, OECD Publishing, Paris.
- OECD (2013) 'Trends in evaluation and assessment', in *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*, OECD Publishing, Paris.
- Pellegrino, J, Chudowsky, N and Glaser, R (2001) *Knowing What Students Know: The Science and Design of Educational Assessment*, National Academy Press, Washington, DC.
- Pellegrino, J and Hilton, M (Eds) (2012) *Education for Life and Work: Developing Transferable Knowledge and Skills in the Twenty-First Century*, National Research Council, Washington, DC.
- Scoular, C, Ramalingam, D, Duckworth, D and Heard, J (2020) *Assessment of General Capabilities: Skills for the 21st-Century Learner. Final Report*, Australian Council for Educational Research, Melbourne.
- Siarova, H, Sternadel, D and Mašidlauskaitė, R (2017) *Assessment Practices for 21st Century Learning: Review of Evidence, NESET II Report*, Publications Office of the European Union, Luxembourg.
- Soland, J, Hamilton, L and Stecher, B (2013) *Measuring 21st Century Competencies: Guidance for Educators*, Asia Society, New York.
- Taylor, R, Fadel, C, Kim, H and Care, E (2020) *Competencies for the 21st Century: Jurisdictional Progress*, Brookings Institution, Washington, DC.
- Thurstone, L (1927) 'A law of comparative judgment', *Psychology Review*, **34**, p 273–286.
- Tuckett, A and Field, J (2016) *Factors and Motivations Affecting Attitudes Towards and Propensity to Learn Through the Life Course*, Government Office for Science, UK.
- Vincent-Lancrin, S, González-Sancho, C, Bouckaert, M, de Luca, F, Fernández-Barrerra, M, Jacotin, G, Urgel, J and Vidal, Q (2019) *Fostering Students' Creativity and Critical Thinking: What it Means in School*, OECD Publishing, Paris.
- Voice 21 (2019) *The Oracy Benchmarks*, Voice 21, London.
- von der Embse, N, Jester, D, Roy, D and Post, J (2018) 'Test anxiety effects, predictors and correlates: A 30-year meta-analytic review', *Journal of Affective Disorders*, **227**, p 483–493.
- William, D (2007) 'Once you know what they've learned, what do you do next? Designing curriculum and assessment for growth', in R Lissitz (Ed), *Assessing and Modeling Cognitive Development in Schools*, JAM Press, Maple Grove, MN.
- World Economic Forum (2015) *New Vision for Education: Unlocking the Potential of Technology*, World Economic Forum, Geneva.

Additional reading

Although not cited explicitly in the text, the following were used in preparing this paper and may be of interest to the reader.

Lucas, B and Smith, C (2018) *The Capable Country: Cultivating Capabilities in Australian Education*, Mitchell Institute policy report No 03/2018, Mitchell Institute, Melbourne.

OECD (2017) 'PISA 2015 collaborative problem-solving framework', in *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*, OECD Publishing, Paris.



BILL LUCAS

About the Author

Bill Lucas is Professor of Learning and Director of the Centre for Real-World Learning at the University of Winchester in the UK. Bill is a co-founder of Rethinking Assessment, a coalition of education leaders, employers, researchers and policy makers looking to reform assessment in England.

In 2017 Bill was appointed to chair the strategic advisory group for the PISA 2021 test of Creative Thinking, which draws on the five-dimensional model for creativity that he developed with colleagues in Winchester. In 2019 the OECD published the results of a four-year, eleven-country study into the teaching and assessing of creativity, also stimulated by Bill's research. In the same year Bill co-authored the Durham Commission on Creativity and Education's first report.

Bill advises the Victorian Curriculum and Assessment Authority in Australia on the assessment of critical and creative thinking, and has worked extensively across Australia over the past eight years with a focus on critical and creative thinking. In 2018 his report for the Mitchell Institute, *Capable Country: Cultivating Capabilities in Australian Education*, suggested a route map for Australia to make the general capabilities a reality in all states.

Bill is a prolific writer and has authored more than eighty books and many research reports. His latest books, written with Ellen Spencer, are *Zest for Learning: Developing Curious Learners who Relish Real-World Challenges* and *Teaching Creative Thinking: Developing Learners who Can Generate Ideas and Think Critically*. His acclaimed critique in 2015 of the education system in England, *Educating Ruby: What Our Children Really Need to Learn*, written with Guy Claxton, asks challenging questions about the future direction of schools.

About the Paper

The author argues that in our current approaches to assessment we are 'using the wrong kinds of nets', especially if we are wanting to 'catch' young people's strengths. He discusses issues with the content of school curricula; models of a more global curriculum and lifelong learning; the roles of skills and competencies in learning; and related problems with educational assessment. He revisits the purposes of assessment, explores promising practices from around the world and provides examples of both visible progress and emerging new directions in assessment. He concludes that we need urgently to develop and implement more nuanced, strengths-based, multimodal descriptions of what young people achieve.