Preparing for a Renaissance in Assessment

Peter Hill and Michael Barber

December 2014

PEARSON

Preparing for a Renaissance in Assessment

Peter Hill and Michael Barber

ABOUT PEARSON

Pearson is the world's leading learning company. Our education business combines 150 years of experience in publishing with the latest learning technology and online support. We serve learners of all ages around the globe, employing 45,000 people in more than seventy countries, helping people to learn whatever, whenever and however they choose. Whether it's designing qualifications in the UK, supporting colleges in the US, training school leaders in the Middle East or helping students in China learn English, we aim to help people make progress in their lives through learning.

INTRODUCTION TO THE SERIES

The Chief Education Advisor, Sir Michael Barber, on behalf of Pearson, is commissioning a series of independent, open and practical publications containing new ideas and evidence about what works in education. The publications contribute to the global discussion about education and debate the big 'unanswered' questions in education by focusing on the following eight themes: Learning Science, Knowledge and Skills, Pedagogy and Educator Effectiveness, Measurement and Assessment, Digital and Adaptive Learning, Institutional Improvement, System Reform and Innovation, and Access for All. We hope the series will be useful to policy-makers, educators and all those interested in learning.

CREATIVE COMMONS

ABOUT THE AUTHORS

Dr Peter Hill has held senior positions in education in Australia, the USA and Hong Kong, including as Chief Executive of the Victorian Curriculum and Assessment Board, Chief General Manager of the Department of School Education in Victoria, Australia, Professor of Leadership and Management at the University of Melbourne, Director of Research and Development at the National Center on Education and the Economy in Washington DC, Secretary General of the Hong Kong Examinations and Assessment Authority and Chief Executive of the Australian Curriculum, Assessment and Reporting Authority.

He is currently a consultant advising on system reform in the areas of curriculum, assessment and certification. He has published numerous research articles and co-authored with Michael Fullan and Carmel Crévola the award-winning book, *Breakthrough*, published by Corwin Press.

Sir Michael Barber is a leading authority on education systems and education reform. Over the past two decades his research and advisory work have focused on school improvement, standards and performance; system-wide reform; effective implementation; access, success and funding in higher education; and access and quality in schools in developing countries.

Michael joined Pearson in 2011 as Chief Education Advisor, leading Pearson's worldwide programme of research into education policy and efficacy, advising on and supporting the development of products and services that build on the research findings and playing a particular role in Pearson's strategy for education in the poorest sectors of the world, particularly in fast-growing developing economies.

Prior to Pearson, Michael was a Partner at McKinsey & Company and Head of McKinsey's global education practice. He co-authored two major McKinsey education reports: *How the World's Most Improved School Systems Keep Getting Better* (2010) and *How the World's Best-Performing Schools Come Out on Top* (2007). He is also Distinguished Visiting Fellow at the Harvard Graduate School of Education and holds an honorary doctorate from the University of Exeter.

Michael previously served the UK government as Head of the Prime Minister's Delivery Unit (from 2001 to 2005) and as Chief Adviser to the Secretary of State for Education on School

Standards (from 1997 to 2001). Before joining government he was a professor at the Institute of Education at the University of London. He is the author of several books including *Instruction to Deliver, The Learning Game: Arguments for an Education Revolution* and *How to Do the Impossible: A Guide for Politicians with a Passion for Education*.

Michael has recently been appointed as Chair of the World Economic Global Advisory Forum.

CONTENTS

FOREWORD

Assessment is a very complex topic. As this essay articulates, it is meant to monitor or to measure what students have learnt. For validity and reliability, and to minimise subjectivity, standardised tests are often adopted and marks are awarded, followed by a process in which test scores are converted into grades. The grades are then recognised as measures of students' learning attainment. But what assessment actually means is seldom articulated. Is it a measure of the body of knowledge that a student has acquired, or is it also a measure of other attributes?

Institutes of higher education have often found such assessment grades to be so lacking in substance for admission purposes that many of these institutes have introduced other modes of assessment so as to gauge the other desired attributes of their candidates. The complexity of assessment is further compounded by the way in which test scores are utilised. Apart from being considered for entry into further education, they are also used for the purpose of accountability of schools or the system, as well as the performance of teachers.

In the twentieth century, the standardised test approach could be valid and reliable, though never perfect. However, in the twenty-first-century landscape, where the demands go beyond just knowledge and technical skills, there is, indeed, a need for an 'assessment renaissance' so that the desired attributes can be meaningfully monitored or measured. However, in this new world, where there are so many 'drivers' that are impacting education

systems, a fundamental issue that must be first clearly articulated is 'What is the purpose of education in this new world that we live and work in?' Only when we can articulate with clarity the purpose of education in terms of the learning outcomes that the education process aims to achieve can we then articulate what an assessment renaissance implies so that the 'what' and 'how' of assessment can be crystallised.

For an assessment renaissance to be meaningful, it also needs a total cultural shift within society to accept the different 'what' and 'how' of assessment. The current mindset of assessment is all about test scores, irrespective of whether the meaning of the test scores is well clarified. In realising the outcomes of the assessment renaissance, there may not always be a test score to contend with. It may just be a series of qualitative descriptions of the extent to which a student may have demonstrated various attributes that cannot be quantified. Can society accept such assessment outcomes?

Going forward, assessment will remain a complex issue, no matter what form the assessment renaissance may take. It is here that the importance of research and development into assessment issues cannot be overemphasised. If the 'what' and the 'how' can be conducted with clarity of meaning, and considered valid and reliable with minimal subjectivity, and if society at large can be educated about the need for such a renaissance, then there will be light at the end of the tunnel. I believe this will take

time, but the journey must start immediately. I congratulate the authors for writing this think piece, which sets out so clearly where we have come from and where we need to go.

Professor Lee Sing Kong
Vice President for Education Strategies, Nanyang Technological University Director, National Institute of Education, Nanyang Technological University (2006–14)

EXECUTIVE SUMMARY

SETTING THE SCENE

Three core processes lie at the heart of schooling:

1. curriculum (deciding what students should learn);
2. learning and teaching; and
3. assessment (monitoring student learning).

When well executed, they work together in a symbiotic fashion, and all other activities function in support of this triad. Of the three, this essay focuses primarily on assessment, but we are aware that it is not possible to talk about changes in the field of assessment without relating them to a much wider set of changes taking place in education.

The educational revolution

We believe that two game-changers are at work that will shake the very foundations of the current paradigm of school education. The first is the push of globalisation and new digital technologies, which are sweeping all before them. The second is the pull inherent in the realisation that the current paradigm is not working as well it should any more. Even the top-performing systems in the world have hit a performance ceiling.

Key elements of the education revolution

Table ES.1 summarises what we see as six key changes that characterise this revolution. The seeds of each of these key changes are everywhere to be seen. There are schools

and systems that are already operating in or contemplating moving towards some of the directions indicated, but this is inevitably a sporadic process.

ASSESSMENT: A FIELD IN NEED OF REFORM

The primary purpose of educational assessment is to seek to determine what students know, understand and can do. While that would seem a relatively straightforward intention, in the real world of policy and practice, educational assessment is complex and frequently controversial.

This essay reviews the key purposes of assessment, namely its use in formal assessment programmes for the purposes of certification, selection and accountability; and its formative use in classrooms and schools for improving learning and teaching. We have also sought to illustrate why assessment, when used for these purposes, is so often controversial, difficult and a barrier to change. The key challenges we have highlighted are summarised in Table ES.2, which contrasts what we ideally want from formal assessment programmes with what we typically get.

Table ES.1 Key features of the education revolution.

Key element	Overthrown and repudiated	Replaced by
1. Capacity to learn	Practices reflecting an assumption that students commence school *tabula rasa* and with an innate and fixed capacity to learn and profit from formal education	• Practices that build on prior learning and reflect a belief in the potential for all students to learn and achieve high standards, given high expectations, motivation and sufficient time and support
2. The curriculum	Curricula that emphasise memorisation of unrelated facts and breadth at the expense of depth	• A greater emphasis on deep learning of big ideas and organising principles • More explicit and systematic attention to cross-curricular skills, capabilities, understandings and dispositions that support lifelong learning and living in the Knowledge Society of the twenty-first century
3. Education policy	The school as the focus of educational policy	• The student as the focus of educational policy and concerted attention to personalising learning
4. Opportunity to learn	Current age and time-bound parameters: • age–grade progression • 9.00–4.00 school hours • open 200/365 days a year	• Students able to progress at different rates and with time and support varied to meet individual needs • Significantly increased access to care and education to better align with the realities of modern living and working • Greater use of the home, the community and other settings as contexts for 24/7 learning
5. Teaching	Predominantly teacher/text instruction, with schools and classrooms as the physical and organisational places for all formal learning and with the classroom teacher as the imparter of knowledge	• Increasing reliance on sophisticated tutor/online instruction with greater differentiation in educator roles and the creation of learning partnerships between and among students, teachers and families, with the teacher as the 'activator'
6. Teacher quality	Teaching as largely under-qualified and trained, heavily unionised, bureaucratically controlled 'semi-profession' lacking a framework and a common language to describe and analyse teaching	• Teaching as a true profession with a distinctive knowledge base, a framework for teaching with well-defined common terms for describing and analysing teaching and strict control by the profession itself on entry into the profession

Table ES.2 Assessment: a field in need of reform.

The ideal	The norm
Assessments that can accommodate the full range of student abilities	• Assessments unable to assess accurately at either end of the ability distribution, or away from critical cut-scores • Assessments within tiered credentials or tiered assessments, with resulting problems of cost, logistics, cross-tier comparability and capping of student aspirations
Assessments that provide meaningful information on learning outcomes	• Over-reliance on grades or levels that reveal little about what the student can do • Feedback to schools on student performance typically provided too late and too broad-brush to be of value in improving learning and teaching • Assessments used to generate a single score for each student which is then further summarised at the school or system level as a percentage meeting a nominated cut-score – a volatile statistic, hiding more than it reveals about performance, particularly shifts in performance on either side of the cut-score. Alternatively, summarised as a mean score unadjusted for intake and other characteristics beyond the control of the teacher or school
Assessments that accommodate the full range of valued outcomes	• Tests and examinations dominated by questions assessing low-level cognitive processes and failing to capture such valued outcomes as practical, laboratory and field work, speaking and listening, higher-order cognitive processes and a range of inter- and intra-personal competences (so-called 'twenty-first century skills')
Assessments that support students and teachers in making use of ongoing feedback to personalise instruction and improve learning and teaching	• Assessment policies that pay little or no attention to formative assessment and to providing teachers with the tools and the capacity to use it on a daily basis • An absence of validated learning progressions, efficient processes for collecting and analysing data and easy-to-use assessment tools
Assessments that have integrity and that are used in ways that motivate improvement efforts and minimise opportunities for cheating and 'gaming' the system	• Assessments that carry undue weight in high-stakes decision-making, increasing the risks of cheating and 'gaming' the system

TRANSFORMING ASSESSMENT

This chapter describes ways in which new thinking and new digital technologies are transforming assessment and overcoming current barriers and limitations. We begin by considering how these changes affect formal assessment programmes, such as those used for certification/selection and accountability purposes, and then move to consider assessment as part of the ongoing process of learning and teaching. Finally, we indicate how a better balance between various purposes of assessment and a closer alignment of assessment with curriculum and teaching can be achieved as a result of the radical changes in thinking and practice made possible by these developments.

Transforming formal assessment programmes

Increasingly, formal assessment programmes serving certification, selection and account-ability purposes are being administered online, not only as part of a broad trend within modern society but also, more particularly, because the online assessment environment offers a number of major advantages once the technical problems of access have been addressed. These include:

- assessing the full range of abilities;
- providing meaningful information on learning outcomes;
- assessing the full range of valued outcomes;
- maintaining the integrity of assessments.

Transforming assessment, as part of the ongoing process of learning and teaching

We then consider assessment undertaken at the point of learning, at the teacher–student interface typically (although not necessarily in classrooms), as part of the ongoing process of learning and teaching.

Developers of next-generation learning systems don't start with preconceived notions of any of these components but completely rethink the whole delivery process and how to best assist teachers to connect all of the elements so that they operate seamlessly. We can follow the logic of these systems with the aid of the diagram in Figure ES.1.

Curriculum

Starting at the top of Figure ES.1 is the curriculum, but one looking quite different to curriculum documents of the past, consisting of online interactive multidimensional maps at several different scales that can be interrogated in different ways, depending on one's focus or query.

Assessment

Going clockwise around the diagram, the next element is assessment. Yes, personalised learning systems move straight from the curriculum (deciding what students need to learn) to assessment, because effective learning and teaching require that one begin with the students and their individual starting points.

Resources

In generating instructional sequences, learning tasks and associated assessment activities, next-generation learning systems will embed or search out the resources that most closely match students' learning needs, accessing both purpose-built, commercially available materials and the rapidly expanding collections of public-domain and creative-commons resources.

Data management and analysis

It was not so long ago that almost all information about students and their learning

Figure ES.1 Next-generation learning system.
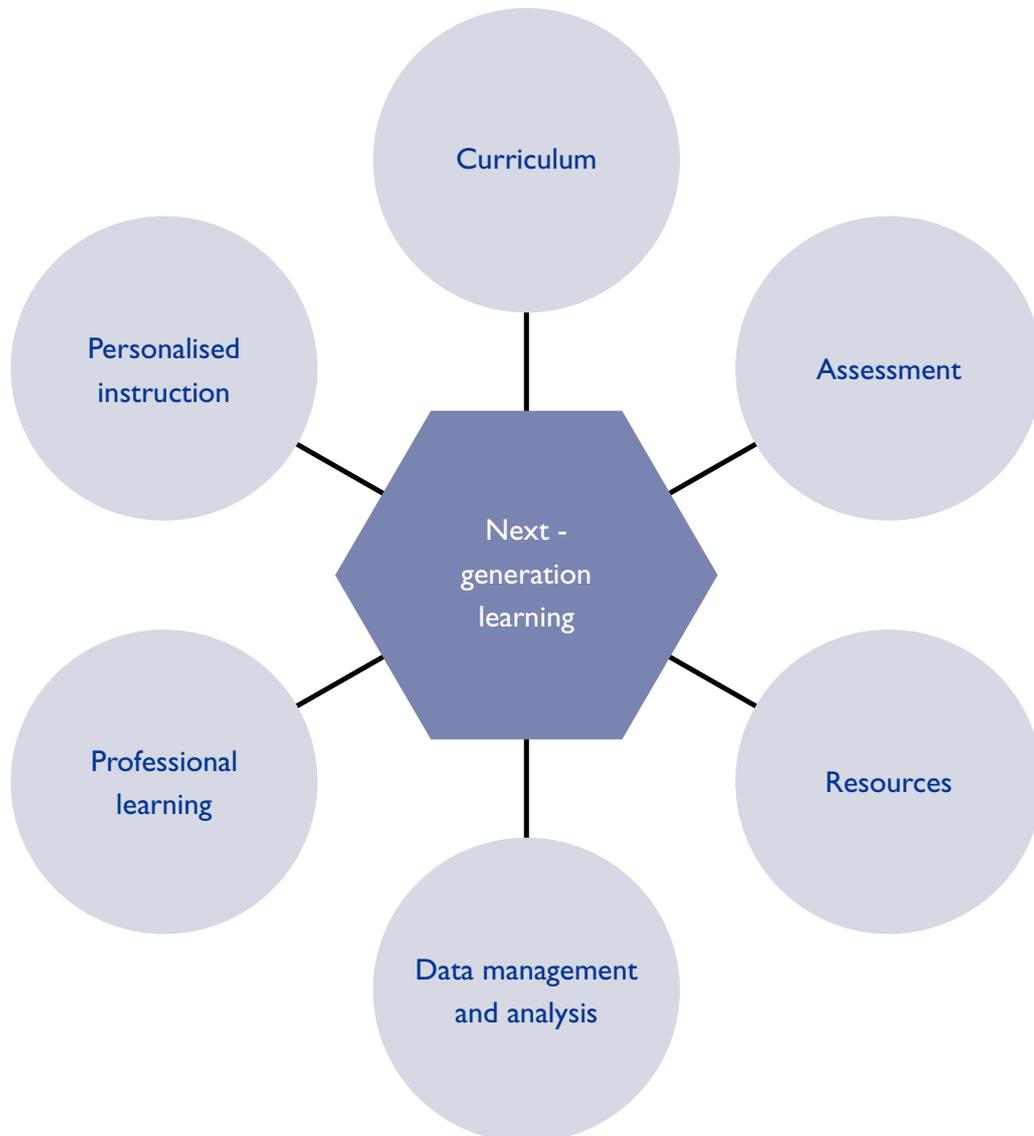
was contained within teachers' books of marks, attendance registers, student record cards and student reports. Next-generation learning systems will create an explosion in data because they track learning and teaching at the individual student and lesson level every day in order to personalise and thus optimise learning. Moreover, they will incorporate algorithms that interrogate assessment data on an ongoing basis and provide instant and detailed feedback into the learning and teaching process.

Professional learning

In next-generation learning systems, the teacher retains the key role in fostering the learning for each student, but the job itself changes. Learning systems of the future will free up teacher time currently spent on preparation, marking and record-keeping

Table ES.3 Transforming assessment.

The ideal	How new thinking and technologies can help
Assessments that can accommodate the full range of student abilities	• Use of adaptive testing to generate more accurate estimates of student abilities across the full range of achievement while reducing testing time

Assessments that provide meaningful information on learning outcomes	• Online environments to facilitate:
	  – the administration of multiple versions of the same test in order to obtain information on performance across a much wider range of the curriculum
	  – the collection and analysis in real time of a wide range of information on multiple aspects of behaviour and proficiency and
	  – more immediate, detailed and meaningful reporting to specific stakeholder groups, such as via smartphone/tablet devices and through the creation of e-portfolios
	• Advances in the application of data analytics and the adoption of new metrics to generate deeper insights into and richer information on learning and teaching

Assessments that accommodate the full range of valued outcomes	• Automated marking to overcome obstacles to the more widespread use of essay and other open-response format questions • Platforms to support the delivery of a new generation of assessments specifically designed to assess deep learning and a range of inter- and intra-personal competences and character traits
Assessments that have integrity and are used in ways that motivate improvement efforts and that minimise opportunities for cheating and 'gaming' the system	• The adoption of (1) more cumulative approaches to approaches to assessment for selection purposes, with opportunities to re-sit; and (2) intelligent accountability systems that utilise multiple indicators of performance, that are designed to incentivise improvement and that avoid the creation of win–lose consequences for stakeholders for outcomes not fully under their control
Assessments that support students and teachers in making use of ongoing feedback to personalise instruction and improve learning and teaching	• Sophisticated online intelligent learning systems to integrate the key components involved in effective instruction and to support a new generation of empowered teachers in reliably assessing a much wider range of outcomes, using instant and powerful feedback on learning and teaching to deliver truly personalised instruction

and allow a greater focus on the professional roles of diagnosis, personalised instruction, scaffolding deep learning, motivation, guidance and care. This is the combination of activities that John Hattie describes as 'teacher as activator' (2009: 17).

Personalised instruction

With all the above in place, it is then possible to talk confidently about personalised instruction, which is the final and most crucial component of Figure ES.1. By personalised instruction, we mean instruction that is adjusted on a daily basis to the readiness of each student and that adapts to each student's specific learning needs, interests and aspirations. The fundamental premises of personalised learning have been a part of the writings of educators for decades but have, in recent years, become a realisable dream, thanks to the advent of new digital technologies.

Rethinking, aligning and rebalancing assessment

In short, new thinking and digital technologies are transforming assessment and overcoming many current barriers and limitations. Table ES.3 summarises what we see as the main features of this transformation.

An integrated, multi-level view of assessment

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.

Rather than focusing on discrete assessment programmes, we would suggest that it is more productive to view assessment as serving distinct data needs at three levels:

1 the teacher–student interface (traditionally the classroom);
2 the school; and
3 the system.

The most important level is the teacher–student interface, because this is where learning takes place and where there is the greatest need for assessment data to enable a truly personalised approach to learning and teaching. We would argue that the other levels and purposes of assessment should be built on the assessment carried out at this level.

The challenge for awarding bodies

In considering the future of assessment for certification purposes, the challenge facing awarding bodies is to work out how they can take greater advantage of new technologies to deliver examinations online and, by so doing, enhance their capacity to:

- assess a wider range of valued outcomes;
- create more authentic assessment tasks;
- more accurately assess the full range of student abilities and speed up the process of marking student responses, including those to extended response questions;
- open up the window of time in which examinations may be taken and work towards the longer-term goal of examinations on demand;
- use the potential of online assessment and developments in psychometric methods to more rigorously maintain standards and constantly benchmark them to ensure that these standards are world-class.

The accountability challenges

Designing an effective accountability system involves clarifying who can and should be held to account for what at each level of the system and establishing accountability arrangements that are reasonable and effective and that promote a shared trust in the system. This means being sure that, as far as possible, accountabilities are within the power of the person or organisation being held to account.

In the school educational context, this typically means holding systems, schools and teachers responsible for:

- student growth or progress rather than purely for absolute levels of performance; and
- doing those things that the evidence shows lead to improved outcomes – not just for achievement of the outcomes themselves, which may be only partly attributable to the specific person or organisation being held to account.

Equally important in the design of accountability systems is the need to take into account capacity-building requirements, particularly those related to teachers' assessment literacy.

A FRAMEWORK FOR ACTION

In this chapter, we propose how policy-makers, schools and school-system leaders and other key players can prepare for an assessment renaissance, ensuring that they maximise the benefits of new developments and changes in thinking while avoiding the potential downsides. We present a framework for action that allows change to be implemented in ways and in timeframes suited to the starting points, capacity and readiness of schools and systems.

1   Think long-term.
2   Build partnerships.
3   Create the infrastructure.
4   Develop teacher capacity.
5   Allow variation in implementation.
6   Adopt a delivery approach.
7   Communicate consistently.
8   Apply the change knowledge.

In conclusion, we see the changes in thinking about assessment as leading to a veritable renaissance – a revival in thinking and practice that promises to overcome many of the key limitations of the current paradigm and to put assessment more fully in the service of both the curriculum and learning and teaching. Governments, systems, schools and those within them all have critical roles to play in bringing this about.

1. SETTING THE SCENE

Three core processes lie at the heart of schooling:

1 curriculum (deciding what students should learn);
2 learning and teaching;
3 assessment (monitoring student learning).

When well executed, the three work together symbiotically, and all other activities function to support this triad. This essay focuses primarily on the third process: assessment. It is often the piece that sits uncomfortably with the other two, and it is the one we believe is currently lagging behind in efforts to secure improved learning outcomes for all.

There is now a growing consensus among leaders in the field that we are on the verge of a radical change in thinking and practice regarding assessment in school education.[1] However, the exact form of this change depends very much on how we anticipate, envision, plan for and shape it.

If this change is managed skilfully, we believe that education will witness an assessment 'renaissance' – a 'rebirth' of the core purposes of assessment – that will lead to a much better alignment of all three processes. More specifically, we see assessment changing in ways that will help secure a floor of high standards for all, removing current achievement ceilings and supporting a focus on those higher-order thinking and interpersonal skills vital for living and learning in the twenty-first century.

In *Preparing for a Renaissance in Assessment*, we seek to:

- summarise the reasons for and the nature of these changes;
- indicate how governments, schools and school-system leaders and other key players can prepare for these changes and ensure they maximise the benefits and avoid potential downsides; and
- provide a framework for action to enable change, which can be implemented in ways and timeframes suited to the starting points, capacity and readiness of schools and systems.

We have sought to avoid going into technical arguments and details but instead to provide a widely accessible and readable overview of the more significant changes without oversimplifying the underlying complexities.

The field of assessment in school education is vast, so we have necessarily been selective. Thus, we have opted to review developments affecting K-12, but with an emphasis on the assessment of fifteen- to eighteen-year-olds. We consider a number of uses of assessment but emphasise high-stakes uses for the purposes of certification, selection, accountability and improving learning and teaching.

[1.] See, for example, Gordon Commission on the Future of Assessment in Education (2013) and Global Education Leaders Program (2014).

As we started to write this essay, we realised that we could not discuss changes in the field of assessment without relating them to a much wider set of revolutionary changes taking place in education. So, in order to understand the 'whys' and 'hows' of the coming renaissance in assessment, we will begin with a brief overview of the more fundamental changes happening more broadly in education, of which assessment is but one vital part.

THE EDUCATIONAL REVOLUTION

Change is a constant in the modern world, and we certainly witness it in education (although when the dust settles we often remark on how the fundamentals seem to stay the same). In many areas of educational policy and practice, we simply see pendulum swings. Every now and then, however, radical change occurs that completely upsets the old ways of doing things. Such change is revolutionary in character since it overthrows and repudiates established methods and replaces them with an entirely new order.

One hesitates to use the term 'revolution' when talking about fundamental changes in education: after all, no parent welcomes the notion of their children being caught up in anything revolutionary. Furthermore, schools have been among the most stable institutions of society and are not prone to radical change.

Looking back, we can see that formal education's basic structures and modes of delivery have barely changed over the past 140 years. That is something one cannot say of health care, public transport or policing.

Despite many recent innovations, schools continue to provide the same kinds of functions and are recognisably similar to what they have long been, consisting of classrooms, halls, libraries, staffrooms and school grounds for recreation and sport. Instruction continues to be delivered by a teacher, who teaches a class of students of the same age, all progressing through a standard curriculum at the same pace, with new teachers each year. Despite considerable experimentation with new arrangements and new technology, rows of tables and chairs and students working with paper, pen and printed texts continue to predominate. The school year and the school day reflect the demands of an agrarian society that has long since disappeared, with teachers and students enjoying long holidays and short hours that are out of alignment with the working days and hours of their parents and guardians, who face challenges in organising child care. In brief, school education has been characterised by constant surface-level change and periodic calls for a thorough overhaul, but the fundamentals have remained surprisingly constant.

So, not for the first time, we need to take stock and ask the question, 'Are we currently witnessing changes that have more fundamental and far-reaching consequences and that will lead to a reconceptualisation of school education?' We have concluded, as have many other commentators, that this time things are different. In particular, we believe that two game-changers are at work that will shake the very foundations of the current paradigm of school education. The first is the push of globalisation and new digital technologies, which are sweeping all before them. As Hannon and colleagues observe, this is an argument that has been 'exhaustively rehearsed, but is no less valid for that' (2011: 2). The second is the pull inherent in the realisation that the current paradigm is no longer working as well as it should.

Globalisation: the key driver of revolutionary change

The key force for change in the modern world is, and will continue to be, globalisation in all its manifestations (economic, environmental, political, cultural, social and technological). The big driver for all these changes is technology. Digital technologies, in particular, represent the next, rapidly accelerating phase of human evolution. Those of us who operate daily in the world of Web 2.0 can already envisage the magnitude of changes that schools must undergo – and which are already under way in many places. But we can barely conceive of what life might be like in the predicted scenarios of Web 3.0 and beyond, where unlimited access to the web will have become a right and an affordable necessity, artificial intelligence will have surpassed individual human intelligence in many areas, and the internet may indeed have become conscious.[2]

Digital technologies and the internet are transforming almost all aspects of life and creating what has been called the 'Knowledge Society'. This is characterised by

- universal and instant access to knowledge;
- rapid obsolescence of knowledge and the disappearance of generally longer-term jobs dependent upon old knowledge;
- exponential increase in new knowledge and the creation of generally shorter-term new jobs dependent upon new knowledge; and
- the imperative for ongoing learning to update and connect knowledge.

The new world order brought about by globalisation and the emergence of the Knowledge Society has enormous implications for the work of schools, for how education is provided and, indeed, for the very existence of schools as we currently know them.

Let's consider the purposes of education. In the past, it was possible to talk with some certainty about the kind of education needed to prepare young people for life and work, and with some confidence about the pathways it would open up to various careers. In the new world, there is much less certainty about the sorts of jobs that may be needed in the future or the kinds of challenges daily living might involve.

Whole categories of jobs, which until recently employed large numbers of people, are disappearing. At airports, staffed check-in counters are being replaced by self-serve kiosks; the same thing is happening at supermarkets, where self-service tills are replacing checkout staff. Bank tellers and retail sales staff are being replaced by internet banking and online shopping. Anything that can be automated is being automated. While particularly true of many low-paid, unskilled jobs, this also applies increasingly to 'white-collar' occupations and the professions. At the same time, new jobs are being created, but companies are struggling to recruit people with the relevant skills. For example, as illustrated in Figure 1.1, recent evidence from Eurostat indicates a widening skills gap in digital jobs in the European Union, with demand far outpacing both the actual (current) and projected supply of graduates with relevant mathematical, science and engineering backgrounds (European Commission 2013: 85).

How should we prepare young people for such a world? There are those who argue

---

[2.] See, for example, Heylighten (2012).

Figure 1.1 Actual and projected development in digital jobs in the EU: vacancy and graduate numbers.
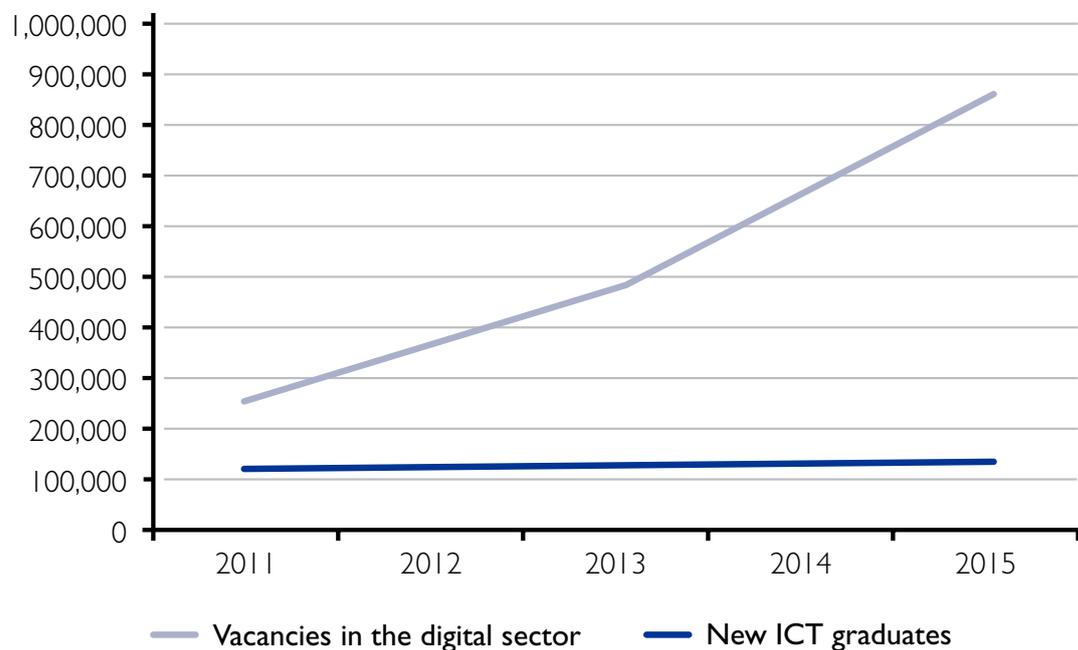
that knowledge of the fundamentals of the disciplines that have long formed the core of traditional school subjects remains vital. At the same time, there are those who call for:

• less emphasis on memorisation of unrelated facts and a greater emphasis on deep learning of big ideas and organising principles (the least obsolescent aspects of knowledge);

• more explicit and systematic attention to a set of skills, capabilities, understandings and dispositions that run right across the traditional subject-based curricula and that facilitate response to change and the rapid acquisition of new knowledge;

• a greater emphasis on 'doing' in addition to the acquisition of knowledge and on allowing living, learning and action to come together in our conceptions of the educated person.

We agree with these points and don't believe they are in conflict.

Discussing knowledge of the core disciplines, Daniel Willingham has observed (2006: 1):

research literature from cognitive science shows that knowledge does much more than just help students hone their thinking skills: it actually makes learning easier. Knowledge is not only cumulative, it grows exponentially. Those with a rich base of factual knowledge find it easier to learn more – the rich get richer. In addition, factual knowledge enhances cognitive processes like problem solving and reasoning. The richer the knowledge base, the more smoothly and effectively these cognitive processes – the very ones that teachers target – operate. So, the more knowledge students accumulate, the smarter they become.

In other words, what we are really asking for is more. Yes, we need to be careful to avoid an overloaded curriculum. Yes, we must ensure there is space for deeper learning of the more important content, which does imply acquiring a rich base of factual knowledge and, beyond that, the ability to understand and apply it. But yes, we also want to ensure, in a more systematic, conscious and explicit way, that, as students learn in specific areas of the curriculum, they are also acquiring key cross-curricular skills, capabilities and dispositions through direct engagement with a curriculum that blends living, learning and action. A number of systems have undertaken major revisions of curricula to address the need to reduce content coverage in order to promote deeper learning, with Singapore one of the first to take decisive action (Ng 2008).

Embedding so-called 'twenty-first-century skills' or 'next-generation learning' into the curriculum has proved much more challenging. These learning outcomes are increasingly seen as critical to equip young people with the skills required to be ongoing learners who can navigate an ever-changing world of work and find fulfilment in their lives. Learning outcomes include the well-understood basics of literacy and numeracy but also involve an education characterised by deep learning and the ability to think, learn, inquire, problem-solve, create, relate and also to manage oneself and one's learning.

Discussion of these higher-order thinking, inter- and intra-personal skills has often taken place without any real agreement on meanings and definitions, and with little research evidence of their importance or even whether they can be taught successfully. The publication of the report of the Committee on Defining Deeper Learning and 21st Century Skills represents a significant step towards clarifying the fundamental definition and research-related questions (see Pellegrino et al. 2012).

In addition, progress has been made on scoping and sequencing these skills or competencies within the context of the overall curriculum. For example, the online Australian Curriculum for K-10 students gives prominence to seven general 'capabilities':

1   literacy;
2   numeracy;
3   information and communication technology capability;
4   critical and creative thinking;
5   personal and social capability;
6   ethical understanding;
7   intercultural understanding.[3]

Each has been scoped in terms of the key outcomes relevant to each capability and sequenced into six levels spanning years K-10. Examples are given, with hyperlinks to specific content areas within mainstream curriculum subjects where these capabilities are particularly relevant and can be developed.

However, the task is not one of simply adding a new set of 'skills' to the curriculum but of continually challenging our concepts of what it means to be an educated person. Here, again, it is a matter of more, not less. In addition to knowledge of the disciplines and cross-curricular skills and understandings, schools are being expected to provide young people with an appreciation of, and engagement with, the big challenges of the modern world, such as sustainability, peace and conflict, the widening

---

[3.] See http://www.australiancurriculum.edu.au/generalcapabilities/overview/general-capabilities-in-the-australian-curriculum (accessed 18 November 2014).

gap between rich and poor, population and resources.[4] In other words, schools are expected to prepare young people to be informed and actively engaged citizens.[5]

One example of where this has been taken seriously is Hong Kong's new credential for students at the end of Year 12, the Diploma of Secondary Education, which requires all students to study, in addition to Chinese language, English language, mathematics and between two and four other subjects of their choosing, a subject called 'Liberal Studies'. The aim is to ensure that all students develop an understanding of the major issues confronting society in the twenty-first century and are equipped with the critical thinking skills they need to make informed, critical judgements about these issues.

Beyond skills or competencies and new understandings, there are calls for schools to pay more attention to developing the character traits and dispositions in young people that will support them in confronting the unprecedented changes taking place in the world around them, such as resilience, adaptability, entrepreneurialism, sensitivity to cultural and personal differences and the disposition to think and act ethically. Cultivating such outcomes is quite a different matter to imparting skills and understandings, because it means engaging students in situations where these qualities matter and can be experienced, reflected upon and nurtured.

Whatever name we give to the disparate set of learning outcomes that constitute next-generation learning, it is clear that they are central to education in the twenty-first century. While many of these outcomes have

long been recognised as important, they have often fallen outside the scope of what has been mandated, made explicit, assessed or certificated. As a consequence, it has been all too easy for them to remain at the level of rhetoric rather than at that of deliberate policy.

> ' New models of learning and teaching are evolving that make traditional classroom, teacher and textbook modes of formal learning obsolete '

Globalisation and the new technologies have fundamental implications, not only for what students need to know and be able to do but also for how it will be taught. Thanks to high-speed internet access, the low cost of devices such as smartphones and tablet personal computers, social media and the evolution of the semantic web, users can find, share and combine information more easily. New models of learning and teaching are evolving that make traditional classroom, teacher and textbook modes of formal learning obsolete.

Some form of 'blended learning', in which a part of what students learn is through online delivery of content and instruction with elements of personalisation for when, where and at what pace, is increasingly becoming the norm, although the form it takes varies enormously, as does the quality.

But deeper, technology-enabled transform-ations are on the horizon. Big publishing and information technology companies, in conjunction with universities and foundations, are embarking on the design of new, fully integrated online learning systems that use detailed learning progressions and continuous monitoring of progress and responses to

---

[4.] A comprehensive framework for considering fifteen global challenges of the early twenty-first century has been developed by the Millennium Project. See http://www.millennium-project.org/millennium/challeng.html (accessed 15 November 2014).
[5.] Regarding the importance of education for citizenship, see in particular, Feith (2011).

deliver finely calibrated instruction that reflects students' learning styles, needs and aspirations. A key motivation behind the development of these more 'personalised' learning systems is the expectation that they will make learning more engaging and more efficient. It is hoped, too, that they will accelerate progress for students who have fallen behind. They have significant implications for the role of teachers, especially their knowledge and skillset.

Glimpses into the future can be had now in pioneering schools across the world. Significantly, the new digital technologies are not just an option for advanced economies, they also offer affordable options for countries in the developing world, particularly through the use of mobile phones (m-learning) to reach places where there are no schools, teachers or libraries.

In summary, the increasing availability of powerful and transformative interactive digital technologies is redefining how learning takes place in schools and all other settings. They are key ingredients of the education revolution.

The performance ceiling

Digital technologies and the new Knowledge Society that they are creating, of themselves, would probably be sufficient to fuel the education revolution, but, as we indicated earlier, there is another game-changer at work, namely the 'pull' factor inherent in the growing realisation that the current paradigm of school education is no longer working as it should.

For many advanced nations, there are clear indications from longitudinal surveys of achievement that a performance ceiling has been reached in the delivery of learning outcomes and in closing achievement gaps. Investment in school education is no longer yielding the returns it once did, when the focus was on access rather than outcomes.

In the USA, which has extensive longitudinal data on performance, NAEP (National Assessment of Educational Progress) survey results indicate that overall performance has improved very little since the 1970s.[6]

But the USA is not alone. Figure 1.2 shows annualised changes in performance in reading and mathematics across PISA (Programme for International Student Assessment) assessments for the top nine countries between the first survey results (either 2000 or 2003) and the most recent 2012 survey. (The error bars are 95-per-cent confidence intervals around each change score.) In the case of reading, only two of the top nine performing countries in the first survey (Japan and Korea) recorded a statistically significant improvement, and in the case of mathematics, none did. This was despite significant efforts and additional resources directed at improving outcomes in each of these countries.

In addition, some of the high-performing countries (notably Australia, New Zealand and Finland) have experienced a statistically significant decline in performance levels rather than an improvement. In short, patterns of results from longitudinal surveys of achievement such as NAEP and PISA would suggest that there are limits as to how much more productivity can be squeezed out of school systems operating within the current paradigm.[7]

[6.] For a commentary on this phenomenon, see Tucker (2013b).

[7.] It should be noted, however, that there are those who argue that tests such as PISA, which seek to provide a common yardstick across nations, are not sensitive to improvements in teaching and learning. PISA does not assess how well students have learned a specific curriculum but rather their ability to apply understandings in reading, mathematics and science to everyday problems and situations.

Figure 1.2 Annualised change across PISA assessments of reading and mathematics for top nine performing countries. Source OECD (2013b).

Readings

Mathematics
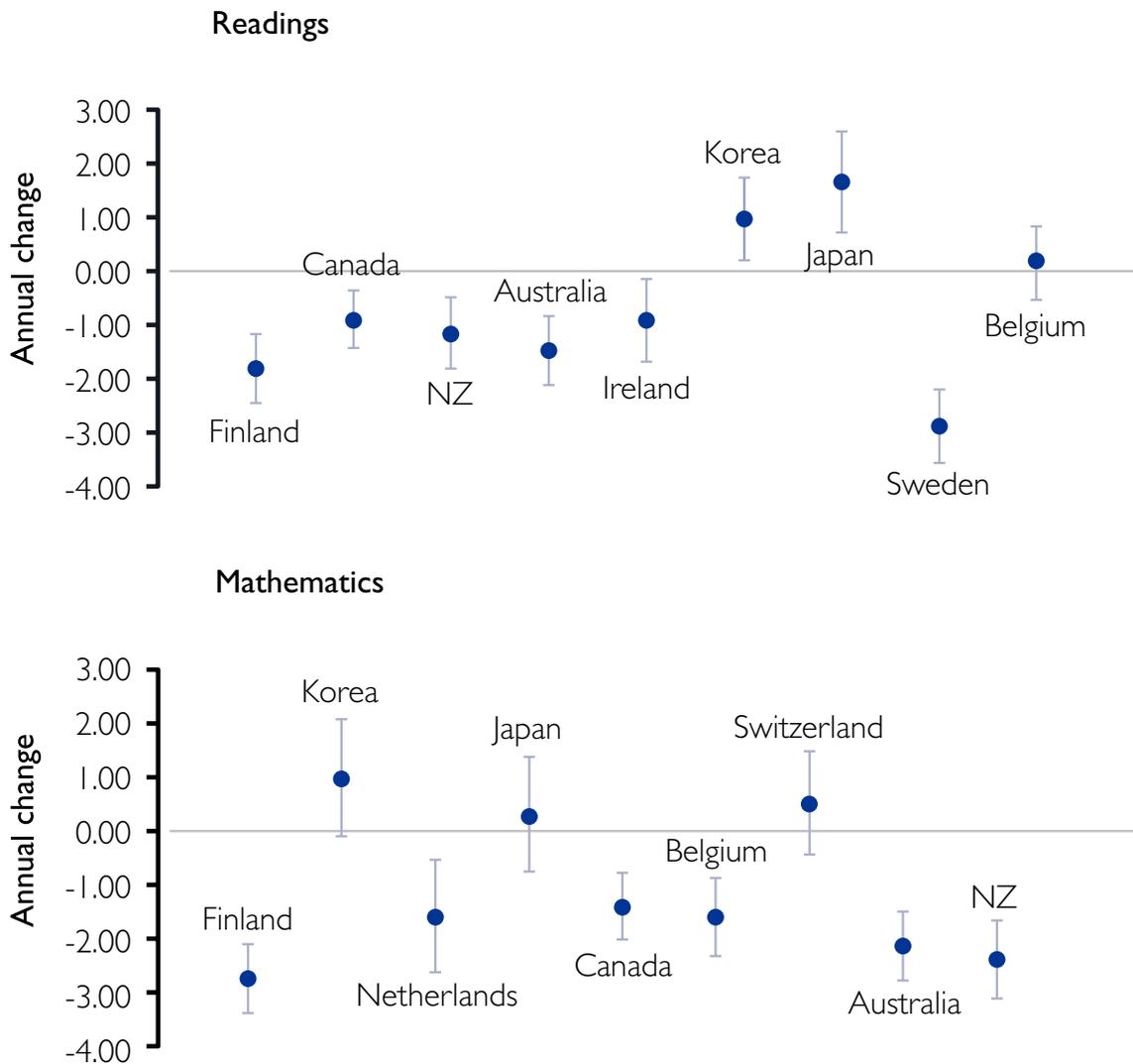
Much of the attention given to improving learning outcomes has been directed at the school level. Analyses of the 2009 PISA data indicate that in the participating countries, after adjustments for demographic and socio-economic characteristics, around 20 per cent of the variance in reading performance could be attributed to differences between schools (OECD 2011: Table IV.2.2a). The same analyses of 2012 data indicated that around 15 per cent of the variance in mathematics performance could be attributed to differences between schools (OECD 2013c: Table IV.1.12a). In other words, there are substantial differences between schools even when their intake characteristics have been taken into account. Research into school effectiveness, much of which was undertaken in the 1980s and early

---

7. It should be noted, however, that there are those who argue that tests such as PISA, which seek to provide a common yardstick across nations, are not sensitive to improvements in teaching and learning. PISA does not assess how well students have learned a specific curriculum but rather their ability to apply understandings in reading, mathematics and science to everyday problems and situations.

1990s, has provided us with a good knowledge of the more powerful school-level levers for improvement. Strong educational leadership, a small number of strategic priorities and a climate of high expectations of student behaviour and learning are among the factors that have delivered remarkable and rapid turnarounds.

However, estimates of school effects can be misleading. Analyses that take into account the fact that students are not only taught within a given school but are also in a particular class within that school, result in much lower estimates of the variance in outcomes at the school level but high proportions of variance at the class level. For example, in one such study conducted by Hill and Rowe in Australia in the 1990s, it was found that fitting a two-level model (students within schools) to local assessment data resulted in estimates of school effects of 17.6 per cent for English and 16.6 per cent for mathematics (very similar to the OECD two-level model outcomes). However, three-level modelling (students within classes, within schools) resulted in estimates of 8.2 per cent for English and 5.4 per cent for mathematics at school level, but 43.7 per cent for English and 56.4 per cent for mathematics at class level (Hill and Rowe 1996).

In other words, it matters more which class a student is assigned to than which school they attend. This is not an altogether surprising conclusion when one considers that learning takes place in classrooms with a specific teacher and a class of students with particular backgrounds, but it points to the fact that, in order to improve learning, it is important to focus on what is happening in individual classrooms and on the quality of teaching received by each student.

‘ quality of teaching is the key to unlocking significant improvements in outcomes ’

There is now a wide consensus that quality of teaching is the key to unlocking significant improvements in outcomes. In 2007, Barber and Mourshed, in *How the World's Best-Performing School Systems Come Out on Top*, concluded that three things matter most:

1 getting the right people to become teachers;
2 developing them into effective instructors; and
3 ensuring that the system is able to deliver the best possible instruction for every child.

In response to the call for a greater focus on teaching quality, many nations have initiated work on clarifying teacher roles and expectations, improving the quality of recruits into teaching, ensuring that pre-service teacher training includes a solid foundation of professional practice and systematically building opportunities to reflect on and enhance their practice into teachers' daily lives. In a few countries, but particularly in the USA, a key part of the solution is seen as the implementation of systems of teacher accountability for student learning, with direct links between individual teachers and their students' test scores.

However, a succession of other commentators, beginning with Dan Lortie in 1975 and most recently Jal Mehta (2013), have reached a more fundamental conclusion.[8] They believe that, in many nations, improvements to the quality of teaching can only come through the transformation of teaching from a largely

---

[8.] Lortie is quoted in the insightful and scholarly review of the field by Grossman and McDonald (2008).

under-qualified and trained, heavily unionised, bureaucratically controlled 'semi-profession' into a true profession with a distinctive knowledge base, a framework for teaching, well defined common terms for describing and analysing teaching at a level of specificity and strict control, by the profession itself, on entry into the profession. Broadly, we agree with this analysis (noting that this characterisation of teaching is less applicable in many Asian countries) and believe that the performance ceiling will remain until the full professionalisation of teaching, in this sense, has become a reality. This is what Michael Barber has called 'informed professionalism' (2014: slide 3).

Whatever the precise contribution of teacher effects (quality of teaching) or the optimum strategies for maximising them, it is unquestionably the case that the greatest proportion of variance in learning outcomes is at student level. Using data from a study by Hauser, Professor Geoff Masters presents a dramatic depiction of the extent of the overlap in performance of more than a quarter of a million mathematics students in different grades in the USA (2013: Fig. 2.3; see Figure 1.3). Much of the overlap seems to be a consequence of the fact that high-achieving students make steady progress, but low-achieving students make very little progress over time.

The phenomenon of wide variations in performance of students of the same age is observed in almost all studies where vertically equated test data (across age grades) are available. These variations indicate that the greatest opportunities for improvement exist at the student level, but, so far, few systems have been able to significantly narrow achievement gaps within grades.

We would suggest that this is in no small part due to the way in which school education

Figure 1.3 Distributions of students' mathematics achievements (Years 2–7, USA, 2003). Source: Masters (2013).
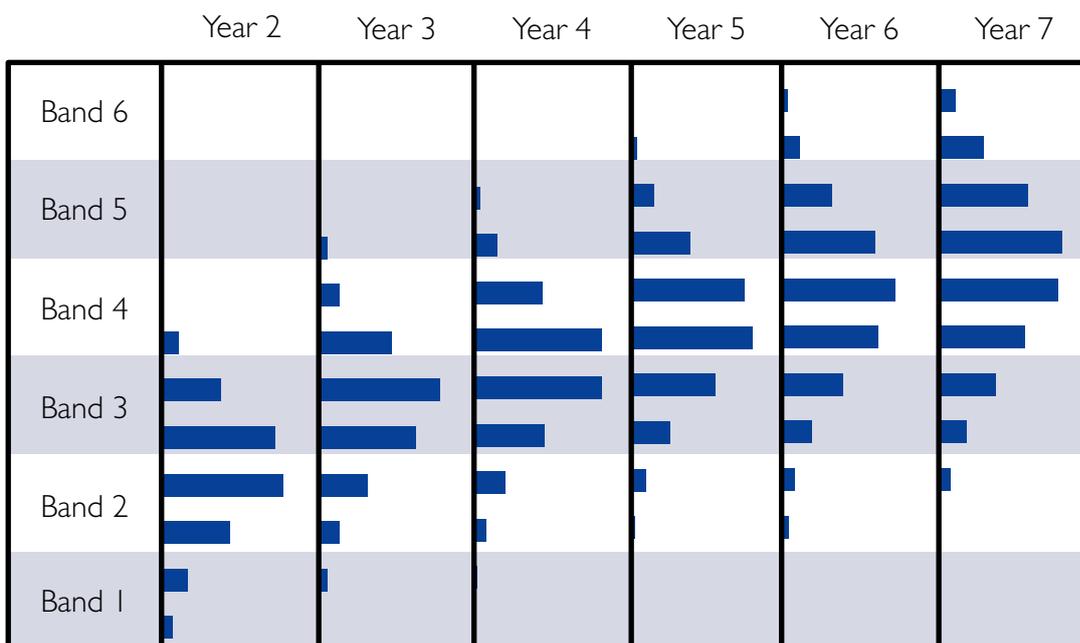
is delivered. The current system has been described as an industrial mass-production model where:

- students are organised into grades, based primarily on age rather than readiness to learn;
- there are discrete curricula and standards for each grade;
- each grade is taught over a single school year by the same teacher; and
- almost all students move to the next grade, new curriculum objectives and standards and new teachers, regardless of how well they mastered the objectives of the preceding grade.

This is a model that could only be effective if one assumed equal starting points and equal readiness to learn. In the real world, this is highly improbable.

The age–grade progression model made sense when it was first invented as a means of educating the masses for a world in which most work required low levels of education and automation had not begun to take over routine tasks. The system efficiently filtered out those who were not able to succeed and directed them to early employment while giving continued access to more and better quality education to the successful few, enabling them to move into professions requiring high levels of education.

It was developed at a time when the accepted view was that the ability to learn and to profit from education was a fixed characteristic of individuals and when students arrived at school with relatively little exposure to formal knowledge. We now have a more positive set of beliefs and understandings about human learning capacity and know through direct experience, supported by research from a number of fields (particularly cognitive science), that potentially everyone can achieve high standards when expectations are high and when the individual is motivated to learn and given sufficient time and support to succeed.[9] In addition, students increasingly come to school having already had significant exposure and access to knowledge, courtesy of television and the internet.

The age–grade progression model is a barrier to realising the new goal of high standards for all because its very structure has an inbuilt assumption of equal time and support for each student. It was never designed to deal with the wide variation in readiness to learn, or to educate all to high standards, or to equip students to live and work in the Knowledge Society of the twenty-first century. It has thwarted at least a decade of intensive reform efforts that have delivered, at best, only the most meagre returns (Fullan et al. 2006).

Instead of putting schools at the centre of improvement efforts, the new paradigm starts with individual students, taking their starting points, motivations and readiness to learn and working back from those to design what is needed to deliver truly personalised learning (Leadbeater 2002). It makes the assumption that 'systems capable of achieving universally high standards are those that can personalise the programme of learning and progression offered to the needs and motivations of each learner' (OECD 2008: 4). In the process, current conceptions of learning and teaching, and of the school itself as the place in which formal education takes place, are being challenged.

---

9. See, for example, Dweck (2006).

KEY ELEMENTS OF THE EDUCATION REVOLUTION

Our thesis, then, is that the 'push' factor of globalisation and the 'pull' factor of the performance ceiling are together giving rise to an educational revolution in which certain long-held beliefs and ways of doing things are being repudiated and replaced by a new set of beliefs and practices. Table 1.1 summarises what we see as six key changes that characterise this revolution. The seeds of each of these key changes can be seen all around us. There are schools and systems that are already operating in or contemplating moving towards some of the directions indicated, but this is inevitably a slow process, and it is likely that the full extent of the transformation brought about by the education revolution will not become evident for some years.

The first key change concerns views about human capacity to learn and profit from formal education. As we have noted, there has been a turnaround in thinking, from an old set of beliefs that saw students as coming to school with an innate and fixed capacity to learn, to a belief in the potential for all to learn and achieve high standards, given high expectations, motivation and sufficient time and support. While the thinking has changed, practice lags behind, and most teachers are required to operate within structures based on outmoded views of human learning capacity and assumptions about prior learning experiences that limit learning opportunities.

The second key change, which concerns the curriculum and what students need to learn, is already well under way and involves a move away from curricula that try to cover too much. Instead, they have a greater emphasis on the deeper understanding of big ideas and organising principles of the disciplines and a more explicit and systematic attention to cross-curricular skills, capabilities, understandings and dispositions to lifelong learning and living in the Knowledge Society of the twenty-first century. There is wide acceptance of the need to move in this direction, but much remains at the level of aspiration rather than reality.

The third key change involves a shift in the focus of educational policy from the school to the individual student and what needs to be done to personalise learning, break through the performance ceiling and enable all to reach high standards. It is a shift that may involve rethinking 'school' as the physical entity in which learning takes place and being more ready to accept the home, the community and other settings as contexts for 24/7 learning.

The fourth key change concerns the opportunity to learn and a repudiation of the age–grade progression model and current historical conceptions of the school day and year, in favour of more open access and provision and with instruction aligned to students' readiness to learn.

The fifth key change concerns how students will learn and involves a movement away from predominantly teacher/text instruction towards an online learning environment in a range of settings, supported by small-group and one-on-one tutorial assistance. Sophisticated educational software will carry much of the burden in delivering authentic twenty-first-century curriculum content, allowing accurate assessment of students' learning needs and interests, tailoring of instruction to the individual student, ongoing evaluation of learning and instruction and delivery of high-quality interactive instructional materials with access to the world's best educators and

Table 1.1 Key features of the education revolution.

Key element	Overthrown and repudiated	Replaced by
1. Capacity to learn	Practices reflecting an assumption that students commence school *tabula rasa* and with an innate and fixed capacity to learn and profit from formal education	• Practices that build on prior learning and reflect a belief in the potential for all students to learn and achieve high standards, given high expectations, motivation and sufficient time and support
2. The curriculum	Curricula that emphasise memorisation of unrelated facts and breadth at the expense of depth	• A greater emphasis on deep learning of big ideas and organising principles • More explicit and systematic attention to cross-curricular skills, capabilities, understandings and dispositions that support lifelong learning and living in the Knowledge Society of the twenty-first century
3. Education policy	The school as the focus of educational policy	• The student as the focus of educational policy and concerted attention to personalising learning
4. Opportunity to learn	Current age and time-bound parameters: • age–grade progression • 9.00–4.00 school hours • open 200/365 days a year	• Students able to progress at different rates and with time and support varied to meet individual needs • Significantly increased access to care and education to better align with the realities of modern living and working • Greater use of the home, the community and other settings as contexts for 24/7 learning
5. Teaching	Predominantly teacher/text instruction, with schools and classrooms as the physical and organisational places for all formal learning and with the classroom teacher as the imparter of knowledge	• Increasing reliance on sophisticated tutor/online instruction with greater differentiation in educator roles and the creation of learning partnerships between and among students, teachers and families, with the teacher as the 'activator'
6. Teacher quality	Teaching as largely under-qualified and trained, heavily unionised, bureaucratically controlled 'semi-profession' lacking a framework and a common language to describe and analyse teaching	• Teaching as a true profession with a distinctive knowledge base, a framework for teaching with well-defined common terms for describing and analysing teaching and strict control by the profession itself on entry into the profession

innovators. What is more, this will be far from impersonal and will provide for increased person-to-person interaction, guidance, instruction and networking. Educator roles will become more differentiated, with a new class of professionals providing high-quality care, direction, guidance, coaching, motivation and management of individual student learning and development. Teachers will focus less on being providers of knowledge and more on assisting students to apply their knowledge, enabling them to overcome barriers to progress and helping them to discern what is important and true.

The sixth and final key change involves the gradual emergence of teaching as a true profession with a distinctive knowledge base, a framework for teaching with well-defined common terms for describing and analysing teaching at a level of specificity and strict control by the profession itself on entry into the profession. This last change is likely to be closely linked to the aforementioned changes in how students learn in the future and to the new roles that educators in schools will perform.

WHEN WILL THE 'REVOLUTION' HAPPEN, AND HOW?

As we have suggested, the education revolution has already begun, but we know from the history of other social revolutions and from the system transformation literature that it is likely to manifest itself first at the fringes and among the more progressive, and that it will have a zigzag trajectory, with some setbacks, failures of nerve and entrenched resistance to change in certain quarters. We also know that there are specific challenges in bringing about change in education as a result of the 'communication gap' that characterises

schooling. As Barber et al. observed in *Oceans of Innovation* (2012: 58),

> *The challenge is that while education reformers are seeking to design a system for 20 years ahead, teachers struggle with the present and parents remember the system of 20 years ago: the conceptual gap is therefore 40 years – a major communications challenge which governments and educators often underestimate. You could argue that the gap is even bigger than this, given that school students of today will still be part of the global workforce 50 years from now.*

Certainly, an enterprise such as school education cannot and should not be changed lightly or in ways that generate confusion and disarray. Change needs to be managed carefully. At the same time, the stakes are high, and the underlying forces for fundamental change are compelling and irresistible. We do no favours to future generations if we do not respond to these changes with the urgency required.

While it would be profitable to continue to explore further the education revolution, our primary focus here is on assessment. We hope that the above discussion has provided a context that makes it easier to appreciate the significance of the radical change in assessment thinking and practice that leading authorities are heralding. It is a radical change that we hope will facilitate broader change in what we want for our young people.

2. ASSESSMENT: A FIELD IN NEED OF REFORM

'Assessment', when used in an educational context, is a broad term referring to 'any appraisal (or judgement or evaluation) of a student's work or performance' (Sadler 1989: 120). It can be done informally, through direct observation and questioning, or more systematically, through the use of rubrics to analyse performance, including classroom activities and tests, or it can be done formally, through system-wide testing programmes and public examinations. In principle, virtually any educational outcome is assessable, although not all can or need to be measured with the same power.

The primary purpose of educational assessment is to seek to determine what students know, understand and can do. While that would seem a relatively straightforward intention, in the real world of policy and practice, educational assessment is complex and frequently controversial. In a recent review of the field, Professor Geoff Masters, CEO of the Australian Council for Educational Research, an organisation that played a leading role in the implementation of OECD's PISA programme, states (2013: 1–2):

> *The field of educational assessment is currently divided and in disarray. Fault lines fragment the field into differing, and often competing philosophies, methods and approaches. … The resulting dichotomies have become the default basis for conceptualising and describing the field: quantitative versus qualitative;*

> *formative versus summative; norm-referenced versus criterion/standards-referenced; tests versus assessments; internal versus external; continuous versus terminal; measurement versus judgement; assessment of learning versus assessment for learning; and so on.*

' tensions arise when assessments designed for one purpose are assumed to be fit for another or when the impact of a secondary use of assessment on core instructional activities is ignored '

Professor Paul Newton pointed out some years ago that much of the confusion and division in the field of educational assessment is not caused by the assessments themselves but by the uses to which they are put. In particular, tensions arise when assessments designed for one purpose are assumed to be fit for another or when the impact of a secondary use of assessment on core instructional activities is ignored (Newton 2007). Newton provided comments on a non-exhaustive list of more than a dozen uses, each supporting a particular set of decisions and having different assessment design implications, and illustrated how readily disarray can arise in the field of assessment when important distinctions are ignored and false dichotomies are perpetuated.

In order to better understand the significance of the radical changes in thinking and practice

on assessment that we and others have foreshadowed, this chapter:

- reviews some key purposes of assessment, including its use in formal programmes for the purposes of certification, selection and accountability and its formative use in classrooms and schools for improving learning and teaching;
- identifies why assessment, when used for these purposes, has often been controversial, difficult and a barrier to change.

ASSESSMENT FOR CERTIFICATION AND SELECTION PURPOSES

In the school education context, the primary purpose of certification is to attest to a student's educational attainments in individual subjects or areas or across a whole programme of study. Certification is typically carried out on completion of high school, although in many systems (such as the UK, Bangladesh, India, Indonesia, Pakistan, Singapore and Thailand), it continues to be a two-step process. Here, the first set of examinations in several subjects is taken at the end of the period of junior secondary education (usually the tenth or eleventh year of schooling) and the second two years later, in a smaller number of subjects studied in depth.

The selection function involves the use of assessment information by admissions staff and employers choosing applicants for positions. This often entails manipulating information generated by the certification process and sometimes supplementing it with further information, including the outcomes of interviews, evidence of achievements, participation in other relevant activities

and referrals or testimonials. The use of certification for selection purposes has high-stakes consequences for students, and, in some countries, where results are used for accountability purposes, for teachers, school leaders and schools too.

The certification/selection functions of educational assessment have a very long and interconnected history. It could be claimed that their origins lie in the national system of examinations created for the purpose of selection into the Chinese Imperial Civil Service some 1,300 years ago. It was the Chinese who invented written examinations based on a set curriculum, leading to the award of degrees and used explicitly for the purposes of selection by merit – principles not taken up in the West until more than a millennium later.

As for the certification/selection of students at the end of their secondary education, the German and Finnish *Abitur*, can be traced back to Prussian law introduced in 1788. The French *Baccalaureate* was created in 1808 under Napoleon. The British Higher School Certificate Examinations (the forerunner of the present-day GCE A-Level examinations) were established in 1918.

All of these examination systems were conceived initially for the purpose of selection into university. They continue to serve this function today, but in a very different context of expanded access and retention, as well as the more general purposes of certification of performance, high-school graduation and selection, regardless of whether students proceed to university, work or other forms of education and training.

In the USA, the use of examinations to certify and select students can be traced back to the New York state legislature's creation of the Regent's examination system. These high-school, end-of-course exams were first administered after the Civil War in 1878. Twenty-three US states run graduation/exit examinations that require a certain standard of attainment in order to receive a high-school diploma. In most states, these high-school examinations are first taken in the tenth grade although students typically complete high school at the end of grade 12.

Selection into universities in the USA has traditionally depended on the use of high-school grade-point averages and scores on standardised scholastic aptitude tests, such as the SAT.[1] The SAT evolved in the 1920s, from the IQ tests developed for the Army during the First World War. Some 1.9 million men were tested on the Army Alpha test of intelligence for 'literates', and the Army Beta test of intelligence for illiterates and non-English speakers, especially new immigrants (Wigdor and Green 1991). These were aptitude rather than attainment tests, associated with the new science of intelligence testing, new theories of psychometrics and the invention of the multiple-choice question, allowing fast and efficient testing of large numbers of candidates. They have had an enormous impact on a wide range of other school testing programmes and, indeed, on the more traditional school curriculum-based examination systems typical of Europe, Australasia and some Asian countries such as India, Pakistan, Malaysia, Hong Kong and Singapore.

In all parts of the world, assessment for certification of students at the end of high school generates ongoing controversy, much of which gets aired annually in the media, while other issues cause internal dilemmas for awarding bodies.

ASSESSMENT FOR ACCOUNTABILITY PURPOSES

Another long-standing use of assessment, and one that has gained huge prominence in recent years, is for the purpose of holding providers (systems, schools and teachers) directly accountable for the performance of their students. In education, as in almost all areas of public and corporate life, ever more complex formal systems of accountability have been created that variously consider compliance with regulations, adherence to professional norms and educational outcomes. It is the last of these which we will focus on here, as it involves a very specific and often controversial use of assessment information.

Making use of assessment information for accountability purposes has a long history. In 1863, the British government, as part of new funding arrangements for elementary education, implemented a system in which funds received by individual schools depended in part on students' performance in examin- ations administered by school inspectors. This system, referred to as 'payment by results', was highly controversial but, nevertheless, a key part of the drive in Victorian England to establish a system of public elementary education for all. This system remained in place for just over thirty years, and, at its height in the 1870s and 1880s, on average around half of the national-level funding an elementary school received depended on the outcome of student examinations. From then on, it was considered inadvisable to use assessment data to hold teachers accountable for student

---

[1.] Most high schools in the USA use a system of five grades in assessing student performance and assign points to these grades as follows: A = 4; B = 3; C = 2; D = 1; F = 0. The average of a student's grade points is referred to as a 'GPA'.

learning. Instead, the emphasis shifted to performance at the school level.

In the UK, the next stage came when the post-Second World War settlement was overthrown by the 1988 Education Reform Act, which at one and the same time introduced market-style reform – devolution of resources to schools, open enrolment and new school models – and sharper accountability, including England's first National Curriculum and national testing of children at ages seven, eleven, fourteen and sixteen.

Implementation of the new assessment arrangements took the best part of a decade, with implementation errors and significant controversy at every step. By 1995, however, national assessment of seven-, eleven- and fourteen-year-olds in mathematics and English (and science for eleven- and fourteen-year-olds) had been introduced. The General Certificate of Secondary Education (GCSE) exam, new in 1988, was reformed and adapted and became the main means of assessment for sixteen-year-olds.

Moreover, by the mid-1990s, transparency had become a major theme of the reforms, and the results of the tests at eleven and fourteen and exams at sixteen and eighteen were published in 'performance tables', which the media promptly turned into rankings.

The Blair government, first elected in 1997, stood by both accountability and transparency, indicating that it would publish more information, including data on a school's progress over time and value-added indicators. Crucially too, its critique was that the previous government had increased the pressure on

schools to perform but had not increased the support to do so. The Blair governments brought major increases in teachers' pay and growth in the numbers of teachers and sought improvements in teacher training and high-quality professional development for all primary teachers in the teaching of mathematics and English.

Importantly, the Blair government argued that only if the system demonstrated its impact, through accountability and transparency, could increased investment in education over the years 1998–2008 be justified, revealing the connection between assessment policy and overall strategy.

In the USA, which has had a long history of assessment for accountability purposes, the No Child Left Behind (NCLB) legislation enacted in January 2002, with cross-party support, introduced what might be regarded as the most ambitious attempt ever to seek to use accountability testing as a means of raising standards.[2] It required states to:

- establish standards for academic proficiency in reading, mathematics and science;
- establish measures for assessing all students in public schools each year in English and mathematics in grades 3–8 and in one of grades 10–12, and later on in science;
- develop a definition of what would constitute 'adequate yearly progress' (AYP) towards the standard that has been set for academic proficiency;
- set targets for schools to enable them to achieve 100 per cent academic proficiency over twelve years; and

---

[2.] No Child Left Behind is a United States Act of Congress that was a reauthorisation of the Elementary and Secondary Education Act.

- set measurable objectives for improved achievement for each of the following subgroups: economically disadvantaged students, students with disabilities and students with limited English proficiency.

In addition, the NCLB legislation incorporated the requirement that states implement 'high-stakes' consequences for schools and districts that failed to demonstrate AYP.

It soon became evident that NCLB targets were unrealisable for many schools. Implementation of the legislation has generated much debate and controversy, with many criticising NCLB for its punitive approach to school accountability and its over-reliance on test scores when making judgements about schools. Without doubt, NCLB made a major contribution to putting achievement gaps firmly on the national agenda. However, no consensus has emerged on how it could be modified or, indeed, whether it should be scrapped in the context of the reauthorisation of the Elementary and Secondary Education Act.

Since 2012, most states have applied for and have been granted waivers from NCLB requirements and, in particular, from exclusive reliance on test scores, in exchange for rigorous and comprehensive plans to 'improve educational outcomes for all students, close achievement gaps, increase equity and improve the quality of instruction in the classroom'.[3] However, recent research indicates that some NCLB waivers allow the flawed accountability practices of the original law to continue and have missed the opportunity to design more effective school accountability systems that might consider non-test-based indicators, student growth, student demographics or

results from subjects other than reading and mathematics (Polokoff et al. 2014).

Common to the UK, the USA and almost all other countries that have adopted accountability testing has been a consensus that outcomes matter, that they should be measured and that schools and systems should be held accountable for them. From a social-democratic perspective, accountability testing has been seen as a way to promote greater equality of opportunity by focusing on groups who have traditionally achieved low educational outcomes and using the data to target interventions. From a neo-liberal perspective, it has been seen as creating an informed public who are better able to exercise choice in where they send their children to school (Hursh 2007), which, in turn, is seen as leading to ongoing improvements in the quality of educational provision as schools compete with one another for students.

Accountability testing certainly resonates with electorates that have come to believe that justice and progress can occur only under conditions of transparency and full knowledge of the facts.[4] Parents believe that they are entitled to know how their child is progressing and how the child's school and school system is performing. They also believe that there is a corresponding entitlement to remediation when their child is not making adequate progress or when the child's school or school system is not performing to expectation.

It is thus no surprise that accountability testing has become common across the world. It may take the form of specially developed standardised tests, particularly to measure basic literacy and numeracy, or use standards-

---

3. See http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html (accessed 18 November 2014).
4. For a compelling discussion of why transparency rules in the modern world, see Fullan (2008).

based external examinations of school subjects originally designed for certification purposes. Evidence is mixed regarding the effectiveness of accountability testing as a policy to improve outcomes. Analyses of PISA 2009 data on factors that might explain differences between countries in student performance revealed that, across OECD countries, the use of standards-based external examinations of school subjects for accountability purposes was associated with higher levels of student performance, but no measurable relationship was found between the prevalence of standardised tests and the performance of school systems (OECD 2010).

In terms of the challenges associated with the use of formal assessment programmes when used for certification, selection and accountability, there are four that have been universal:

1. accommodating the full range of student abilities;
2. providing meaningful information on learning outcomes;
3. assessing the full range of valued outcomes;
4. maintaining the integrity of assessments.

We will discuss each of these in turn.

Accommodating the full range of student abilities

In the case of assessment for certification and assessment purposes, most examination systems were designed initially for the most academically able of the age cohort but have since been modified or redesigned in an attempt to accommodate the expanded range of student aptitudes that have accompanied increased retention rates.

One response has been to offer tiered credentials. For example, in England and Wales, students sitting the GCSE examinations may sit either for Foundation papers, graded G–C, or for Higher papers, graded E–A*, according to their ability and expected performance. Criticism of such arrangements has focused on the potential for tiering to place a cap on the aspirations of students who may have been guided into sitting for lower-tier papers. On the other hand, tiered papers have the advantage of creating a better match between the demands of the assessment and the assumed ability level of candidates and therefore leading to a more efficient assessment. The case for tiering is stronger for subjects such as mathematics and science, which differentiate through the specific content of questions posed, than for subjects such as English and history, which differentiate through the quality of responses to less content-specific questions.[5]

Another response has been to expand the range of subjects available within a mainstream credential, with the intention of better catering for those not suited to or interested in studying traditional academic subjects. But this sets up a hierarchy of esteem among subjects that are manifestly not of equal challenge, even though the subjects in themselves may be equally valuable and worthy of study. If awarding bodies seek to maintain some comparability in the standards of these very different (in terms of demand) subjects, they risk discouraging the very students they wish to encourage. If they decide to award grades that reflect the candidature of each subject, then they generate a problem for users of the credential, particularly for those who require an overall indicator of performance, such as admissions officers.

---

5. For a discussion of tiering in the context of the GCSE, see Oates (2013).

In the case of standardised testing for accountability purposes, an ongoing challenge has been to design tests that can be administered within the time allowed and yet provide accurate measures across the full spectrum of abilities within a given age/grade cohort. As we will indicate in the next section, technical solutions for better assessing the full range of abilities have existed for some time, but test developers have not always been in a position to implement them. The fall-back position has often been to design tests that have maximum reliability around critical cut-scores associated with one or more defined standards of performance, which is perfectly reasonable if what matters is the standard itself. Of course, if one is interested in performance across the full spectrum of abilities, then the number of items and/or score points required to obtain accurate measures rises dramatically. The problem can be appreciated by looking at Figure 2.1, which shows, for PISA 2009 mathematics, both student ability measures and item difficulties on the same scale, allowing the distribution of student ability measures to be compared to the distribution of item difficulties.[6]

It can be seen that the distribution of item difficulties closely follows the distribution of student abilities, which indicates a well-targeted test, but it is also evident that there is only one item appropriate for students in the ability range −3 to −2. To get an accurate estimate of students in this ability range, more items would be needed of matching difficulties.

In brief, tests and examinations are now being required to be more sensitive to performance across a much wider spectrum of student abilities than can be satisfactorily assessed within the confines of a single fixed-item test. As we will see in the next section, however, the problem is now being addressed through various forms of dynamic, adaptive test delivery that can be facilitated by the adoption of onscreen assessment.

Providing meaningful information on learning outcomes

Another big challenge has been that of providing assessment information in ways that are meaningful and facilitate decision-making. This, of course, may have little or nothing to do with the assessments themselves but rather with how assessment information is used.

In the case of assessment for certification, where the primary use has been for selection purposes, many systems have provided some form of ranking statistic, such as a standardised score, a percentile rank or a grade determined by fixed percentages. Normative information can indeed facilitate selection decisions but by itself provides no indication as to what students actually know and can do and can conceal changes in performance levels over time.

As a consequence, most awarding bodies have moved away from normative reporting in favour of a form of standards-referenced reporting in which psychometric methods are used to develop an achievement scale along which cut-scores are identified in order to create a number of hierarchically ordered levels which are then given labels (e.g., grades A–F), accompanied by descriptions of what a typical student achieving a given level/grade is able to do.

[6.] In PISA and most other tests these days, ability measures are estimated on a scale of logits (the logarithm of the probability of correctly answering a question) that typically fall in the range −3 to 3.

Figure 2.1 PISA 2009 mathematics: plot of student abilities and item difficulties. Source: OECD (2012).
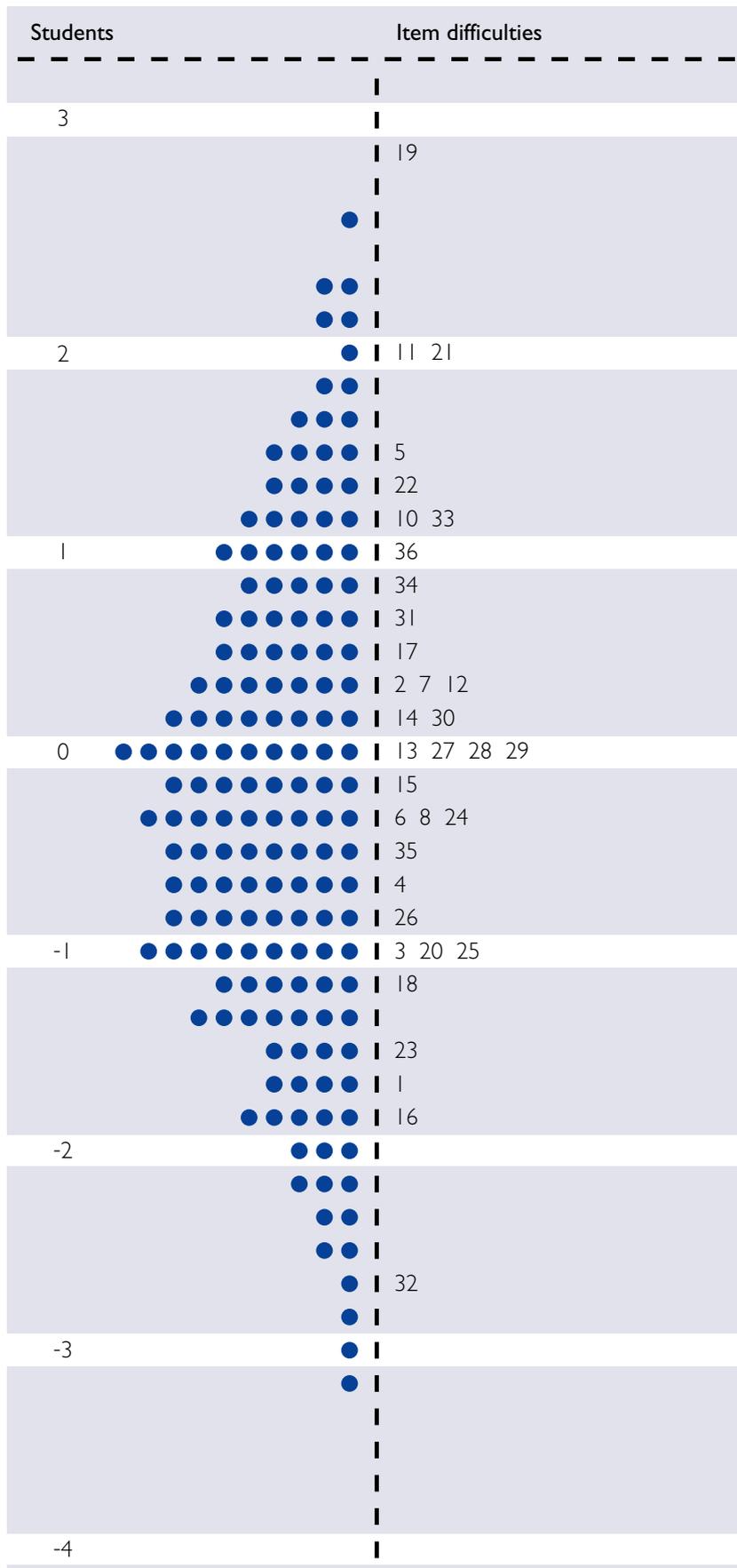
For such standards to have meaning, it is necessary to ensure that they remain comparable over time. For testing situations, the answer is to equate successive tests by embedding a set of 'anchor' test items into the live test. This is routinely done in longitudinal surveys of achievement such as PISA, or literacy and numeracy standardised tests typically administered at state and national levels for accountability purposes.

Ensuring standards are meaningful and remain constant over time is less straightforward in the case of public examinations for which, in the interests of transparency, all examination questions enter the public domain immediately after they are administered and thus cannot be used again for equating purposes. In many parts of the world, reliance is placed on professional judgement to set and maintain standards, including examination of scripts at grade boundaries and comparisons with scripts from previous years. In England, Northern Ireland and Wales, so-called prediction matrices are used to guide examiners in setting boundaries for grades. These matrices make use of students' performance at a previous stage of schooling to predict their performance in GCSE and GCE examinations. Examining bodies have to report and justify significant disparities between predicted grades and actual grades awarded.

> **' the complexity of schooling makes it difficult to capture the performance of a school using simple statistics '**

Challenges of a rather different kind have emerged in providing meaningful information on the outcomes of assessments used primarily for accountability purposes. Policymakers and the public at large seek simple and unambiguous information, such as the percentage of students of a given age/grade meeting a given standard across the system and within a given school. However, the complexity of schooling makes it difficult to capture the performance of a school using simple statistics like this.

To begin with, the information is never as unambiguous as it might seem, thanks to the existence of measurement error, which is a feature of all assessments. This unavoidable fact does not sit well with the average layperson, who typically sees any 'error' as inexcusable and believes all assessments should be completely accurate.

Compounding this problem is the fact that the easy-to-understand 'percent meeting the standard' index is particularly unreliable when it comes to summarising the performance of a school. For small schools, the degree of uncertainty over the percentage of their students meeting a given standard may be greater than the percentage change which the system has declared necessary to demonstrate adequate progress. This unreliability of percentages above a given cut-score statistic also leads to zigzag patterns of performance over time, with some schools erroneously believing that they did very well one year but poorly the next, when in fact the differences may not have been statistically different but rather an artefact of measurement error in the index used.

The 'percent meeting the standard' index has other difficulties.[7] For example, it fails to capture the distribution of performance of the group as a whole and can hide declining performance of the most able students who are well above standard or, conversely, an

---

[7.] For a review of some of these difficulties in the context of the ubiquitous use of the 'percent meeting the standard' index, see Ho (2008).

improvement in the performance of the least able who are well below the standard.

Mean scores are more informative because they take into account the actual scores of each student, but they don't take into account the backgrounds of the students taking the test and other factors beyond the control of the school. Researchers have long advocated greater reliance on so-called 'value-added' measures that seek to adjust for prior achievement, intake and other school and student factors, but there has been a reluctance to embrace them, first because of a commitment to the notion that all should be assessed against the same standard, and second because value-added indices are inherently complex and difficult to grasp for those lacking an understanding of the underlying statistical manipulations.

A claim that has often been made about accountability-assessment programmes is that, in addition to providing information of general public interest, they provide schools and teachers with valuable information for guiding and improving learning and teaching. In other words, an important rationale for administering the tests is that the feedback they provide can enhance teachers' professional practice and give pointers on where to focus school improvement efforts. Often, schools and teachers are given access to detailed breakdowns of the performance of different groups of students on individual test items or on subsets of items assessing specific aspects of the curriculum. Better still, some systems publish detailed analyses of the performances of students on test questions, identifying common difficulties encountered and providing suggestions and identifying resources to teachers on ways in which these can be overcome.

Certainly it is important for schools and teachers to have access to objective information on both the absolute and relative levels of performance of their students. But the potential of test results to improve learning and teaching can be overstated. Results typically reach schools many weeks or even months after students take the tests, by which time they may be in another grade, in another class and with another teacher, so the information is too late to inform practice. Even where there is timely feedback to schools, the information may not be specific or precise enough to inform practice or improve learning in any but a very general way. This is particularly the case in testing programmes in which the test items represent a very broad and light sample from the target domain.

In seeking to use assessments designed for broad system and school accountability purposes to inform daily teaching, it is as well to recall Newton's warnings of the tensions that can arise when assessments designed for one purpose are assumed to be fit for another, or when the impact of a secondary use of assessment on core instructional activities is ignored.

Assessing the full range of valued outcomes

A long-standing challenge in assessment for certification and accountability purposes, and one on which we are now beginning to see significant progress, has been how to assess the full range of valued outcomes. Recent systematic quantitative analyses and benchmarking of curriculum documents with corresponding examination papers have revealed imbalances, with a preponderance of questions relying on relatively low-level cognitive processes such as memorisation, comprehension and problem-solving of a predictable and formulaic nature and few

questions assessing the kinds of thinking skills that result from deep learning.[8]

In some jurisdictions, the traditional essay question, which can often tap into higher-order cognitive skills, has been discarded in favour of multiple-choice or short-response structured question formats, in an effort to improve marking reliability and efficiency and to provide greater access to the full range of student abilities.

There have also been significant gaps in assessment, particularly when it comes to laboratory, field and practical work, oral language and presentations and almost all the inter- and intra-personal skills and competences discussed earlier, which are now seen as vitally important for learning, living and working in the twenty-first century.

Such imbalances and gaps make it impossible to have a complete picture of a student's learning and, more seriously, mean that outcomes not assessed in the examination will receive little or no attention in the classroom. Thus, assessment is dictating and constraining, rather than reflecting, the curriculum.

' assessment is dictating and constraining, rather than reflecting, the curriculum '

The most common response to this dilemma has been to design a system that includes a component conducted at school level to assess outcomes not readily assessed under examination conditions. In order to ensure comparability in standards when these assessments are conducted, various forms of 'moderation' have been devised, including bringing teachers together to review samples of student work, training teachers, inspection

and re-marking of samples by external examiners, tightly defining the nature of the assessment and how it will be scored and statistical moderation using results on the examination paper as the moderating variable.

Faced with the costs of an effective system of moderation, pressure from teachers to relieve them of the burdens that such systems often impose and the difficulties of managing widespread distrust in the integrity of such assessments, many awarding bodies have felt obliged to eliminate the use of school-based assessment or to restrict it only to those instances where it is deemed absolutely essential (such as in the case of orals to assess second-language acquisition). Such trends run counter to the directions emerging in the development of modern curricula that will prepare students for a globalised world and life within the emerging Knowledge Society.

The problem of assessing only a limited range of valued outcomes is, of course, even more acute in most accountability-testing programmes, which typically assess only a small part of the intended curriculum (literacy and numeracy and sometimes core science concepts). In the past, the arguments for focusing on these key areas were unassailable, as outcomes such as literacy and numeracy underpin learning across the curriculum and in later life. But while literacy and numeracy are clearly vital, they are insufficient preparation for life in the modern world. More is being required, and accountability programmes run the risk of missing out on some of the very outcomes that will underpin success in the future.

Finally, there are particular issues associated with narrowly focused accountability-testing programmes that arise from their impact

---

8. An example of research that has investigated the level of cognitive demand in examinations is Clesham (2013).

on other forms of assessment. The Gordon Commission summarised this problem in the context of the USA as follows (2013: 7–8):

> *[assessment] has been seen by policymakers as a means of enforcing accountability for the performance of teachers and schools. … Accountability is not the problem. The problem is that other purposes of assessment, such as providing instructionally relevant feedback to teachers and students, get lost when the sole goal of states is to use them to obtain an estimate of how much students have learned in the course of a year.*

Avoiding this danger calls for a rethink not just of what should be assessed within accountability programmes but also of the fundamental premises underpinning them. This is something we will return to in more detail in the next section, where we identify some of the more promising developments under way to enable assessment of a wider range of learning outcomes.

Maintaining the integrity of assessments

Those responsible for running examinations have always had to cope with the threats posed by those (typically the small minority) who seek to beat the system. As Steven Levitt and Stephen Dubner cleverly illustrated in their best-seller *Freakonomics* (2007), if the incentive is there, some people will do what it takes to get what they want, so perhaps we should not be surprised that people will do all they can to exploit the vulnerabilities of examination systems.

Cheating and corruption were a notorious and well-documented problem throughout the long history of the Chinese Imperial Examinations,

including bribery, paying someone else to sit one's examination (identity fraud) and cribbing (concealing notes). In the modern digital age, smartphones and purpose-built concealed microelectronic devices which can communicate with an outside collaborator or post exam questions live on social-networking websites have introduced a whole new level of complexity and challenge to the task of maintaining the integrity of examinations. This integrity must be maintained without negatively affecting the validity of the examination, infringing on individuals' liberties or otherwise causing undue expense, personal stress or inconvenience to all. Of course, cheating and corruption enter into many aspects of everyday life, so it is no surprise that they should enter into assessment for certification and selection purposes. On the other hand, any system that allows such behaviour to become widespread will inevitably fall into complete disrepute, so this issue needs the closest attention.

If maintaining the integrity of assessments is a challenge in assessment for certification purposes, where problems tend to involve isolated students, it is perhaps an even more serious challenge in accountability testing, in which the stakes are often high for teachers, principals and system officials. Assessments can be compromised by behaviours ranging from excessive drilling to the test to more serious but much rarer instances of professional misconduct. Moreover, the line between right and wrong is not always clear-cut, at least to some. For example, there are documented cases in which a school administrator who had deliberately altered students' responses to give them higher scores declared this behaviour morally defensible as it guarded against potential closure of the local school and its attendant consequences.

In both the USA and the UK, there is evidence that improvements in the performance of schools and students as assessed through high-stakes testing programmes is typically higher than that indicated by performance on parallel low-stakes programmes, giving credence to the view that test-based accountability improvements in learning outcomes reflect, in part, drilling to the test and various strategies to 'game the system'.[9]

This suggests a problem that goes well beyond isolated cases of cheating or manipulating outcomes, and which has little to do with concerns over the nature of the assessments used in accountability testing. Instead, it relates to a clash in values and to underlying faults in

> ' a clash in values and … underlying faults in the accountability arrangements … generate widespread attempts to game the system '

the accountability arrangements that generate widespread attempts to game the system. It indicates a need for some rethinking of the nature of accountability and how it can be built in to systems in ways that are embraced and owned by those being held accountable and that reward those who act in a professionally accountable way in matters reasonably considered to be under their control.

A question we address in the next two chapters is whether challenges in maintaining the integrity of assessment can be ameliorated as part of new developments within the field. But first we turn to the fourth of the key purposes of assessment, namely its formative use in classrooms and schools for improving learning and teaching.

ASSESSMENT FOR IMPROVING LEARNING AND TEACHING

As an integral part of the three core processes referred to at the beginning of the first chapter, the most critical role of assessment is that of monitoring student progress. This provides feedback, which can inform decisions about what to teach next (the curriculum) and provide evidence of the outcomes of learning and teaching. This feedback is most powerful when used by students to adjust their learning strategies and by teachers to make daily, micro-level adjustments to their teaching. When used to inform, guide and personalise learning and teaching, this is known as 'formative' assessment (Popham 2008).

> ' assessment, when used formatively, is one of the most powerful interventions found in the educational research literature '

Through meta-analysis (statistical summarisation) of thousands of research studies, we know that assessment, when used formatively, is one of the most powerful interventions found in the educational research literature (Black and Wiliam 1998; Hattie and Timperley 2007). Despite considerable interest in exploiting its potential, educational policy-makers have struggled to promote it. In the first place, reorienting teachers' professional practice is no easy task and is not something that can be done quickly or without massive and consistent support and encouragement. Second, even when the will has been there, teachers have found it almost impossible to sustain on a daily basis within current models of provision and support, and, as a consequence, it remains underused and its potential unrealised.

---

9. See, for example, Chadowsky and Chadowsky (2010) and Statistics Commission (2005).

When we visit the doctor, we are in a one-on-one situation, and we receive individual attention. If unsure of the diagnosis or treatment, our doctor refers us for further tests or to a specialist. This is routine practice in healthcare. (That is not to suggest that the more personalised approach in healthcare always results in accurate diagnoses but rather that it has a much greater likelihood of doing so.)

When we go to school, we join a class of twenty-five or more students assigned to a teacher who is expected to be able to cope with all but the most extreme learning or behavioural difficulties. Most assessment is informal, unsystematic and takes two forms: (1) ongoing observations of and reflections on students at work; and (2) the posing of questions to monitor responses to instruction. When teachers do assess more systematically, it is invariably for the purpose of making judgements and generating evidence to support a final set of assessment grades. These then appear on students' end-of-term or end-of-year report cards and may subsequently be used for various internal guidance and selection purposes.

To tap fully into the power of formative assessment, particularly for the more critical parts of the curriculum (such as learning to read), it is necessary for teachers to:

- have a clear notion about which aspect or qualities of learning they wish students to develop, in the form of validated 'maps' of the sequence in which students typically learn a given curriculum outcome (variously known as 'learning progressions' or 'critical learning instructional paths' [Fullan et al. 2006: 54]);

- have a simple and efficient process for real-time collection, storage and analysis of large amounts of data about their students;
- monitor students and their progress on a daily basis using a set of structured observations and assessment tools linked to the objectives of each lesson and integrated into learning activities to minimise interruption to normal classroom routines;
- use the data as a starting point for both immediate and longer-term planning and adjustment of instruction explicitly linked to curriculum objectives and tailored to the needs of individual students.

> ' formative assessment: too onerous for the majority of teachers to implement and sustain '

Much of the above has simply not been available, and this has made formative assessment too onerous for the majority of teachers to implement and sustain. But without such a systematic, data-driven approach to instruction, teaching remains an imprecise and somewhat idiosyncratic process that is too dependent on the personal intuition and competence of individual teachers.

This may sound a brutal claim and is certainly not meant as an attack on teachers but rather of the paradigm within which they operate and the impossibility of personalising learning given current conceptions and practices. The issue to be explored in the next chapter is the extent to which new thinking and new digital technologies can remove many of the barriers to full adoption of formative assessment.

Table 2.1 Assessment: a field in need of reform.

The ideal	The norm
Assessments that can accommodate the full range of student abilities	• Assessments unable to assess accurately at either end of the ability distribution, or away from critical cut-scores. • Assessments within tiered credentials or tiered assessments, with resulting problems of cost, logistics, cross-tier comparability and capping of student aspirations
Assessments that provide meaningful information on learning outcomes	• Over-reliance on grades or levels that reveal little about what the student can do • Feedback to schools on student performance typically provided too late and too broad-brush to be of value in improving learning and teaching • Assessments used to generate a single score for each student which is then further summarised at the school or system level as a percentage meeting a nominated cut-score – a volatile statistic, hiding more than it reveals about performance, particularly shifts in performance on either side of the cut-score. Alternatively, summarised as a mean score unadjusted for intake and other characteristics beyond the control of the teacher or school
Assessments that accommodate the full range of valued outcomes	• Tests and examinations dominated by questions assessing low-level cognitive processes and failing to capture such valued outcomes as practical, laboratory and field work, speaking and listening, higher-order cognitive processes and a range of inter- and intra-personal competences (so-called 'twenty-first century skills')
Assessments that support students and teachers in making use of ongoing feedback to personalise instruction and improve learning and teaching	• Assessment policies that pay little or no attention to formative assessment and to providing teachers with the tools and the capacity to use it on a daily basis • An absence of validated learning progressions, efficient processes for collecting and analysing data and easy-to-use assessment tools
Assessments that have integrity and that are used in ways that motivate improvement efforts and minimise opportunities for cheating and 'gaming' the system	• Assessments that carry undue weight in high-stakes decision-making, increasing the risks of cheating and 'gaming' the system

ASSESSMENT AS THE LAGGING FACTOR

In this chapter, we have reviewed the key purposes of assessment, namely its uses in formal assessment programmes for the purposes of certification, selection and accountability and its formative use in classrooms and schools for improving learning and teaching. We have also sought to illustrate why assessment, when used for these purposes, is so often controversial, difficult and a barrier to change. The key challenges we have highlighted are summarised in Table 2.1, which contrasts what we ideally want from formal assessment programmes with what we typically get.

> ' assessment is the lagging factor in providing quality information about learning and teaching and in reflecting the educational needs of students living in the modern world '

The problems identified in Table 2.1 are by no means new. They give substance to Geoff Masters' assertion, quoted at the beginning of this chapter, that assessment, as a field of endeavour, suffers from divisions, controversies and a host of unhelpful dichotomies. They explain why there is now a growing belief that assessment is the lagging factor in providing quality information about learning and teaching and in reflecting the educational needs of students living in the modern world. Assessment has not kept up with new thinking regarding what is important for students to learn, or about how to teach the curriculum, and has all too often been seen as out of alignment with these other two processes that together form the core of learning in schools. There have been some promising steps in the right direction, but in most cases

they represent incremental reforms. From East to West, there is now a growing consensus over the urgent need for assessment systems to align with the fundamental reforms taking place in other areas of school education.

However, there are changes in thinking and new developments that could enable breakthroughs in each of the above challenges. Considered individually, they can be seen as enhancements to the status quo, but collectively they have the capacity to bring about the assessment 'renaissance' we foreshadowed at the start of this essay.

3. TRANSFORMING ASSESSMENT

Let's briefly review what has been suggested so far. At its core, school education is about deciding what students need to learn (the curriculum), about learning and teaching and about assessment (monitoring student progress). Of the three, assessment is the lagging factor and often sits uncomfortably with the other two, for the reasons we have just identified, many of which have to do not with the assessments themselves but with the uses to which they are put.

However, we are on the verge of an education revolution as a result of irresistible external pressures generated by globalisation, new digital technologies and the emergence of the Knowledge Society. Added to which, there are internal pressures in many high-performing countries brought about by a performance ceiling in terms of the improvements in learning outcomes that can be delivered within the current paradigm of school education.

> ' an assessment renaissance … will help secure a floor of high standards for all, remove current achievement ceilings and support a focus on those higher-order thinking and inter- and intra-personal skills vital for living and learning in the twenty-first century '

Assessment is a key part of the coming education revolution. We believe that the possibility now exists to bring about an assessment renaissance that will help secure a floor of high standards for all, remove current achievement ceilings and support a focus on

those higher-order thinking and inter- and intra-personal skills vital for living and learning in the twenty-first century. In this chapter we outline the key elements of these changes.

As we noted earlier, this future is in many respects already with us and can be viewed at the margins of current practice (which is so often where one encounters the new) or is being created by bringing together components that already exist but which have never before been made to work together.

This chapter describes ways in which new thinking and new digital technologies are transforming assessment and overcoming current barriers and limitations. We begin by considering how these changes affect formal assessment programmes, such as those used for certification/selection and accountability, and then go on to consider assessment as part of the ongoing process of learning and teaching. Finally, we indicate how a better balance between the various purposes of assessment and a closer alignment of assessment with curriculum and teaching can be achieved as a result of the radical changes in thinking and practice made possible by these developments.

TRANSFORMING FORMAL ASSESSMENT PROGRAMMES

Increasingly, formal assessment programmes serving certification, selection and accountability purposes are being administered online as part of a broad trend within modern society, but more particularly because the

online assessment environment offers a number of major advances once the technical problems of access have been addressed and the reluctance to abandon 'tried and tested' traditional approaches has been overcome.[1]

Assessing the full range of abilities

We referred earlier to the dilemma of examiners and test constructors in assessing the full range of abilities in many assessment contexts. Test developers find it difficult, if not impossible, to design paper-based examinations and standardised tests that can be administered within the limited time allowed and yet provide accurate measures across the full spectrum of abilities for a given age/grade cohort. Many tests have both floor and ceiling effects. (There are insufficient items to properly assess the highest and lowest achieving students.)

One response has been to develop tiered credentials, while another has been to design tests that maximise reliability around cut-scores associated with one or more defined standards of performance, while accepting greater imprecision of measurement above and below these cut-scores.

Yet another approach, and one that has been known about for decades, involves the use of computer adaptive testing (CAT) and the application of psychometric methods to calibrate a 'bank' of questions of known difficulty. If students perform well on an item of intermediate difficulty, they are presented with a more difficult question. If they perform poorly, a simpler question is presented. Testing proceeds until an estimate of sufficient precision is achieved, which, for most students,

will require many fewer items than had they sat a standard, fixed-item test.

Implementing CAT requires significant upfront and ongoing investment in the required infrastructure, particularly for schools in providing computers and online access, but also in item development, maintenance and the creation of sophisticated software to deliver valid, individually tailored tests while ensuring the accuracy and comparability of ability estimates. Moreover, its use is confined to assessment tasks that can be scored in real time, making it unsuitable for assessing a range of outcomes, including certain higher-order cognitive skills.

A number of states in the USA have implemented CAT programmes, although their use has been constrained by requirements that accountability testing should assess only grade-specific content. Only one state, Oregon, has thus far implemented a CAT system that is part of state accountability arrangements and aligned with grade-level content standards.

In the future, more states will adopt CAT. The Smarter Balanced Assessment Consortium, one of the two state-led consortia working to develop 'next-generation' assessments aligned to the Common Core State Standards (CCSS), is making use of CAT and a bank of more than 21,000 items to deliver online, high-stakes accountability tests.[2]

In the case of public examinations, a further major impediment to CAT is the requirement that all items be released into the public domain after the examination is concluded. Doing so would compromise the integrity of any CAT-

---

[1.] For a wide-ranging, in-depth review of the potential for computers to impact on assessment, see, in particular, the collection of papers in Lissitz and Jiao (2012).

[2.] See http://www.smarterbalanced.org/smarter-balanced-assessments/computer-adaptive-testing/ (accessed 15 November 2014).

based approach or impose unsupportable development costs in annually replacing all items after they have been published.

Significant adaptability can nonetheless be achieved within formal online assessment programmes by adopting other forms of dynamic, multi-stage test delivery such as designing the assessment as a series of 'testlets' or small tests, as indicated diagrammatically in Figure 3.1.

In this particular illustrative design, all students answer the questions in testlet A, and, depending on their performance, they are directed to either B or C. At the end of completing one of these two testlets, they are then directed to one of D, E or F.

Testlets B–C and D–F all contain items that overlap with adjacent tests. Student responses to testlets A, B and C are scored in real time by the computer, but responses to testlets D, E and F may involve open-ended response questions that can be scored by trained professionals at the conclusion of the testing.

This method of creating an adaptive test minimises the number of questions that need to be developed in order to achieve a predetermined level of accuracy, thus making it feasible to release them into the public domain at the end of the testing period – something that would be more problematic with a large item bank in which the questions were intended to be reused.

Considerable research has been undertaken into developing feasible solutions to the problem of obtaining accurate estimates of the abilities of all students tested while reducing testing time and taking away from students the frustration of having to answer questions that are way too easy or the stress of being confronted with questions that are way too hard. In the longer term, once current limitations have been overcome, there is every likelihood that ways will be found to deliver fully adaptive on-demand assessment, with students sitting tests and examinations tailored to their ability whenever they are ready to do so in a system where assessment is continuous rather than a one-shot opportunity.

Figure 3.1 A simple multi-stage adaptive test design.
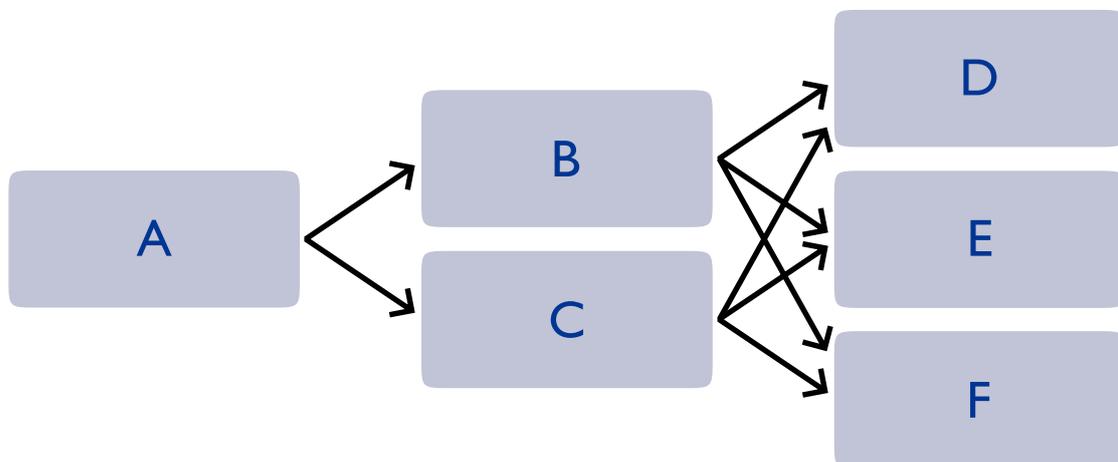
Providing meaningful information on learning outcomes

When discussing some of the limitations of the use of formal testing programmes, we referred to long delays, often amounting to months, in returning results to schools. By the time the data is received, it is often too late to be of much practical value for the students tested and the teachers who taught them.

Online assessment offers the prospect of real-time, instant feedback through automatically scored assessments. What is more, this benefit is not confined to objective and multiple-choice-style questions (something we will discuss in more detail later on).

What this means is that information from formal assessment programmes can be of greater benefit to schools and teachers, who have often been critical of them as not being of great assistance in improving learning and teaching.

Many systems look to formal testing programmes to not only generate an overall picture of performance but also to provide more specific information on performance in discrete areas of the curriculum. However, if a single form of the test is used, the result is a light sampling of the intended curriculum domain. There may be just a single item assessing a particular outcome, making generalised conclusions about performance tenuous.

Better information can be obtained through matrix sampling, whereby students are assigned different forms of the same test. A drawback of matrix sampling is that it entails a certain element of equating error, which can be significant at the individual student level, but the trade-off is that it provides greater information about performance on particular aspects of the curriculum and greater coverage of the curriculum.

An example of a system that has embraced this trade-off is Hong Kong, where there is annual testing of the basic competencies of all students in Grades 3, 6 and 9 in Chinese, English and mathematics. Students are randomly assigned one of three or four versions of the test, thus generating significantly more information about specific outcomes and assessment of a wider range of curricular outcomes. The different forms of the same test are equated so that all students receive an ability estimate that is on the same scale. In addition, sample testing is conducted of students' oral abilities in the two languages. In this way, through multiple forms of the same test and through a sampling approach to harder-to-assess areas, the amount of information about performance across the curriculum is increased significantly.

PISA is an especially good example of an assessment programme that uses matrix sampling to obtain more detailed information on student achievement across the tested domain. In PISA 2012, for example, at least thirteen different test booklets were used in each country, and different forms of the test were randomly allocated to students in a way that ensured that, for each group of thirty-five students, no more than three students received the same test. Through common item equating, it is possible to ensure that, while students take only one of thirteen forms of the test, their scores can be reported on the same scale. (However, PISA does not report results at the student or even the school level, where greater precision in reporting might be required for other purposes.)

Online environments simplify the whole process of administering multiple versions of

the same test in order to improve coverage of the curriculum and to provide more meaningful information on performance across the curriculum.

Currently, most formal assessment programmes focus on generating a single score to summarise attainment. They are conceived within what Robert Mislevy and colleagues (2012: 12–13) refer to as the 'standard assessment paradigm':

> *Data from each student are sparse, typically discrete responses to perhaps 30 to 80 test items. The items are predefined. The target of inference is a student's level of proficiency in a domain framed in trait or behaviourist psychology and defined operationally by the items. Learning during the course of assessment is assumed to be negligible.*

But in an online environment it is possible to not be so constrained, and one might think of assessment as involving 'continuous performances in interactive environments, for example; richer data that encompass many aspects of activity at any level of detail; interest in multiple aspects of proficiency, evoked in different combinations in different situations; learning may occur, and may indeed be an aim of the experience' (Mislevy et al. 2012: 13).

For some purposes, the current paradigm, which involves a predetermined test that seeks to make inference to a single underlying trait, such as 'literacy' or 'mathematics', may continue to make sense, at least for the time being. But there is a price to pay.

Achievement is inherently multidimensional, and, while there will be contexts in which

the overall score is what users want to have, there is growing demand for more accurate knowledge of the specific strengths of students across a range of outcomes. This applies particularly to some of the so-called twenty-first-century skills that are clearly discrete and that do not lend themselves to traditional forms of assessment and reporting.

In the future, we can expect online assessment to collect a wide range of information on multiple dimensions of outcomes, and data analytics to mine far more information from students' responses, thus enabling a more rounded and complete picture of a student's achievements and capabilities. This requires new kinds of assessment, as we have mentioned earlier, but also new kinds of metrics to summarise achievement and performance in those domains that require separate forms of reporting.

Looking even further into the future, more dramatic changes in the ways of assessing and characterising individuals may become possible – ways that personalise the assessment by looking not just at multidimensional aspects of performance but that also take into account the particular situation and context in which individuals were observed and other person-specific information about the performance, challenging the sufficiency of what Mislevy refers to as 'the "one-size-fits-all" presumption of standard assessment, which defines the target of inference in terms of an assessor specified domain of tasks, to be administered, scored, and interpreted in the same way for all students' (2013: 89).

Finally, online environments open up possibilities for more immediate, detailed and meaningful reporting of formal assessment data that is tailored to the needs of specific

stakeholder groups, including parents, teachers, school administrators, employers, tertiary institutions and the general public, using the internet and smartphone/tablet devices. In addition, online environments offer richer ways to record the achievements and significant experiences of individual students, particularly via lifelong personalised student e-portfolios.

For example, the Hong Kong Education Bureau coordinates an online Student Learning Profile system for its high schools, providing a range of online templates that schools can use or adapt to capture information to supplement the Hong Kong Diploma of Secondary Education Examination results, including:

- academic performance in school (other than examination results);
- other learning experiences;
- performance/awards gained outside school; and
- students' 'self-accounts' of their learning experiences and career goal setting.[3]

The system has gained wide acceptance among universities in Hong Kong, mainland China and overseas.

More meaningful information on learning is what assessment reform is ultimately all about, as it is the key to better choices, lifting performance and the motivation to improve. New thinking and new technologies offer the prospect of much progress in the quality of information on students' achievements and capabilities.

Assessing the full range of valued outcomes

As noted earlier, there is evidence that many formal assessment programmes are characterised by a preponderance of questions relying on relatively low-level cognitive processes such as memorisation, comprehension and problem-solving of a predictable and formulaic nature; few questions assess the kinds of thinking skills resulting from deep learning and the capacity to apply what one has learned to new situations; and no questions address the inter-personal and intra-personal skills and competences now seen as vitally important.

To a large extent, this situation reflects an over-reliance on multiple-choice questions which came about thanks to inflated concerns for reliability, at the expense of validity, and by economic considerations over the costs of marking essays and open-ended questions.[4] But it also reflects the absence of established ways to assess these outcomes rigorously.

In some situations, a partial answer may be to both reduce the frequency of testing and to increase the proportion of questions in tests and examinations that assess higher-order cognitive processes. In the USA, the widely adopted CCSS have presented the two assessment consortia charged with developing aligned assessment systems a significant challenge in assessing a range of higher-order cognitive processes and problem-solving capabilities. Sample items published on their respective websites indicate that significant progress has been made in meeting this challenge.[5]

---

3. See http://cd1.edb.hkedcity.net/cd/lwl/ole/SLP/SLP_01_intro_01.asp (accessed 15 November 2014).

4. When questioned recently on their views about the current state of testing in the USA, Howard Everson, Vice President for Research at the College Board, said he thought the importance of reliability had been overblown in the USA. Professor Robert Linn, one of the country's leading assessment experts, agreed, adding that reliability was less important than comparability and validity and fairness. See the full interview at Tucker (2013a).

5. See, for example, http://parcc.pearson.com/sample-items (accessed 15 November 2014); and http://www.smarterbalanced.org/sample-items-and-performance-tasks (accessed 15 November 2014).

For other outcomes, the way forward may be to learn from systems that have succeeded in assessing hard-to-test outcomes through the use of performance assessments. The Colorado Department of Education defines performance assessment as assessment based on observation and judgement. It has two parts: the task itself and the criteria for judging quality. Students complete a task (give a demonstration or create a product), which is evaluated by judging its level of quality using a rubric. Examples of demonstrations include playing a musical instrument, carrying out the steps in a scientific experiment, speaking a foreign language, reading aloud with fluency, repairing an engine or working productively in a group. Examples of products can include writing an essay, producing a work of art, writing a lab report, etc.[6] David Conley and Linda Darling-Hammond describe a number of performance assessments in *Creating Systems of Assessment for Deeper Learning* (2013).

Pearson's Center for NextGen Learning and Assessment has published a Framework of Approaches to Performance Assessment that sets out different approaches to assessing a wide range of valued learning outcomes that are not easily assessed using traditional testing approaches.[7]

Both of these consortia in the USA have developed performance tasks that assess higher-order thinking and problem-solving capabilities, in many cases making use of technology-enhanced item formats and detailed scoring rubrics that require professional judgements of the quality of students' responses. One example (Deer in the Park), developed by PARCC (Partnership for Assessment of Readiness for College and Careers) for its prototyping project, was the fourth-grade sample question shown in Figure 3.2.[8]

Figure 3.2 Sample PARCC fourth-grade question.

The perimeter of the rectangular state park shown is 42 miles.

State Park

8 miles
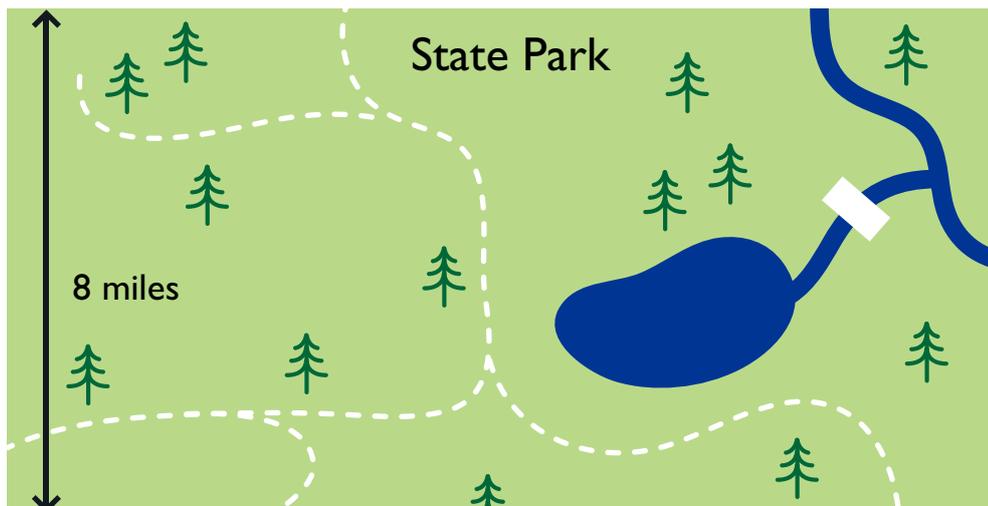
A ranger estimates that there are 9 deer in each square mile of the park.
If this estimate is correct, how many total deer are in the park? Explain your answer using numbers, symbols and words.

---

6. See http://www.cde.state.co.us/contentcollaboratives/phase2performance (accessed 15 November 2014).
7. See http://paframework.csprojecthub.com/?page=home (accessed 15 November 2014).
8. Available at http://www.ccsstoolbox.com (accessed 15 November 2014).

This is clearly a challenging problem for fourth graders, involving the operations of addition, subtraction, multiplication and perhaps also division (up to multi-digit) and requiring knowledge of areas, perimeters, rectangles and how to solve for an unknown in a perimeter. Moreover, students might choose to tackle the problem in different ways and arrive at the correct answer. The accompanying rubrics allow for a possible 6 points of credit for their response.

A barrier to the use of such assessments has been the difficulty and costs of objectively rating open-ended student responses. However, advances in artificial intelligence in combination with online delivery are helping to overcome some of these barriers. While it might at first seem implausible that a machine could mark an essay, several studies have indicated that automated essay-scoring systems employing artificial intelligence are capable of achieving levels of reliability equal to or exceeding that of trained human raters.[9] Some widely used systems include Project Essay, Grader™, Intelligent Essay Assessor™, E-rater®, IntelliMetric™ and Bayesian Essay Test Scoring System™.

All systems developed thus far have certain limitations, but so too does human rating.[10] Currently, automated scoring of extended response questions is usually deployed in high-stakes testing contexts in conjunction with human rating (to provide a second rating or to quality-assure the human ratings, for example). As automated essay-scoring technologies improve, they can be expected to play a much more prominent role.

In the USA, the two federally funded assessment consortia, PARCC and Smarter Balanced, both intend to incorporate automated scoring into their common core state assessments, planned for implementation in 2014. This indicates a growing confidence in automated essay-scoring as means of enabling the assessment of a wider range of outcomes in the context of large-scale, high-stakes testing programmes.

A more fundamental solution lies in using digital technologies to support the adoption of a new generation of assessment tasks specifically designed to assess deep learning and other key outcomes not amenable to assessment via traditional tests and exam-inations. Computerised assessment opens up the prospect of presenting students with tasks that are interactive, that make use of simulations in which students manipulate variables to achieve a desired result, that are dynamic, with the task itself subject to new information and changing circumstances, and that generate a detailed log of students' interactions with the task. Furthermore, it offers solutions to the age-old problems of validity and reliability across those assessing, by allowing not only automated scoring of keyed responses but also rating of a wider range of response types, including performances captured using video and sound recordings, by multiple professional assessors in different locations and at different times.

Jim Soland, Laura Hamilton and Brian Stecher, in 'Measuring 21st-Century Skills: Guidance for Educators' (2013), provide (in addition to a review of the issues involved) interesting case studies of new measures that indicate what is possible right now. One example they highlight,

---

9. See, for example, Dikli (2006).
10. For a comparison of strengths and weaknesses of automated and human scoring, see Zhang (2013).

which does not require overly sophisticated technology, is Mission Skills Assessment, a scientifically based assessment of six character traits – teamwork, creativity, ethics, resilience, curiosity and time management – which has been developed by the Independent Schools Data Exchange and ETS (Educational Testing Services) in the USA. For each trait, an overall assessment is achieved by combining multiple indicators of the relevant construct, including student self-reports, teacher observations and situational judgement tests. In this way, it has proven possible to achieve high levels of reliability (as measured by both internal consistency and test–retest reliability) and of validity (in terms of predicting student academic outcomes).

A more high-tech example is the OECD's proposal for assessing collaborative problem-solving as part of PISA 2015 (OECD 2013a). This will be a fully computer-based assessment in which a student interacts with a simulated collaborator or 'avatar' in order to solve a complex problem.

Both Mission Skills Assessment and PISA's assessment of collaborative problem-solving represent examples of the first tentative steps in the unfolding of next-generation digital assessment.

Maintaining the integrity of assessments

When the stakes for individuals are high, risks to the integrity of assessments will also be high. That is human nature, and something technology cannot change. Accountability is vital, but if it is implemented in ways that provoke fear rather than motivation and the capacity to improve, then the accountability system itself is the problem and should be adjusted.

In the case of assessment for certification and selection purposes, one-shot examinations can place students under great pressure to perform, particularly in some Asian countries where academic expectations are high and failure to excel can cause great loss of face for students and their families. These pressures can be reduced through more cumulative forms of assessment and/or a system in which students have opportunities to take examinations when they are ready to sit them and to re-sit them in order to improve grades.

In the case of assessment for accountability purposes, undue pressures on teachers and school and system administrators can be reduced though the use of multiple indicators of performance, as opposed to exclusive reliance on test scores, and on accountability for implementing policies and practices aimed at improving student progress, as opposed to student attainment data that takes little or no account of the circumstances and influences affecting attainment.

The quality of assessments is also a factor to consider. Questions with one correct answer (such as multiple-choice questions) are particularly vulnerable to cheating, but questions that require higher-order thinking, open responses and demonstration of a student's underlying thinking in arriving at an answer are less vulnerable (assuming one can authenticate that it is the work of the student, perhaps with the help of voice-recognition software, secure browsers and equipment to detect unauthorised use of cellphones and other devices).

That said, new developments in technology can nevertheless be of assistance. One of the greatest fears for administrators of examinations and tests is security prior to administration. Papers

may be stolen or inadvertently mislaid and subsequently distributed or published for all to see, thus invalidating the entire examination and resulting in enormous costs and logistical problems. Online assessment can dramatically reduce the risk of this occurring. It can also eliminate the risk of papers being tampered with after the test or examination has been completed.

Of course, new developments in technology, particularly smartphones and the wider availability of sophisticated hidden listening devices and transmitters, have greatly enhanced the capacity for cheating. The internet is replete with ads brazenly offering access to highly organised cheating services. Systems struggle to deal with these technology-driven forms of cheating and may have to go to uncomfortable lengths to counter it. For example, for the 2012 university entrance examinations in China, bras were reportedly banned as they set off the metal detectors installed to monitor students for listening devices as they entered examination halls (Phillips 2013).

So, while technology may devise ways to prevent cheating and gaming the system, it offers no panacea for many current forms of assessment used for certification, selection and accountability purposes. A better strategy for ensuring the integrity of assessment may be to create the right incentives and avoid win–lose consequences for stakeholders of outcomes not fully under their control. However, intriguingly, the ultimate solution may lie in the potential of a new generation of assessments designed primarily to monitor and inform ongoing learning and teaching, which is what we turn to next.

TRANSFORMING ASSESSMENT AS PART OF THE ONGOING PROCESS OF LEARNING AND TEACHING

Now we move from formal assessment programmes undertaken for certification, selection and accountability purposes to consider assessment undertaken at the point of learning, at the teacher–student interface, typically (although not necessarily) in classrooms, as part of the ongoing process of learning and teaching.

We have referred to the age-old disconnect that is common between assessment and the other two core activities – deciding what students need to learn and teaching the curriculum. We also noted a paradox: when used formatively by students to adjust their learning strategies and by teachers to make daily, micro-level adjustments to their teaching, formative assessment is one of the most powerful interventions known in improving learning outcomes. Yet it is neither widely practised nor, until very recently, given significant attention by education policy-makers and administrators. The reason for this neglect, we suggested, is that within the current model of provision and support provided, it is almost impossible for teachers to sustain formative assessment on a daily basis.

We also referred to the performance ceiling created by the current one-size-fits-all age-grade progression model and the reasons why next-generation learning must be all about differentiating instruction and ensuring that it is optimal for each and every student.

There is now the prospect of tackling these limitations head-on with the development of sophisticated online intelligent learning systems (or 'ecosystems') that facilitate

the integration of these core activities. Dr Ramona Pierson, one of the leaders in the development of new software to drive more personalised approaches to learning and teaching, summarises the challenge as follows (2011: 1):

> *The goal of the Learning Ecosystem (LE) is to bring critical resources into the hands of teachers to transform the teaching and learning moment. By leveraging a fully integrated learning ecosystem, education will finally be able to fulfil the goal of developing a mass customised, personal learning solution at scale for all students and educators.*

Most of us are familiar with the way in which so much of the content of learning and teaching that formerly existed in print form (such as curricula, lesson plans, student and teacher texts and resources, assessments and teacher professional-development materials) has migrated online in recent years. Developers of next-generation learning systems such as Pearson don't start with preconceived notions of any of these components but completely rethink the whole delivery process and how to best assist teachers to connect all of the elements so that they operate seamlessly.

We can follow the logic of these systems with the aid of the diagram in Figure 3.3.

Curriculum

At the top of the diagram is the curriculum, but one looking quite different to curriculum documents of the past, consisting of online interactive multidimensional maps at several different scales that can be interrogated in different ways, depending on one's focus or query.

At the largest scale, one might view the entire curriculum in broad detail. At the smallest scale, it could be a small segment of the curriculum, broken down into a sequence of step-by-step items of skill and knowledge required in order to attain more generalised curriculum outcomes. These are what are known as 'student learning progressions' and are the basic units on which learning ecosystems are built (Popham 2008: 83). They are much more granular than one finds in most articulations of curriculum or core standards and, in the context of learning ecosystems, are not static but are continually refined on the basis of system feedback on how students are learning.

> **' In next-generation learning systems, the teacher can construct and deconstruct the curriculum in ways uniquely relevant to students '**

In addition, at each scale, one would be able to view the curriculum according to one's particular focus. In next-generation learning systems, the teacher can construct and deconstruct the curriculum in ways uniquely relevant to students, building upon local curriculum standards and content and supplemented with other content, but always within a common framework and using a consistent set of terminology and codes, allowing easy identification and cross-referencing. In this way, they will be able to connect more readily with students' interests and aspirations and engage them more deeply in the learning.

Assessment

Going clockwise around Figure 3.3, the next element is assessment. Personalised learning systems move straight from the curriculum (deciding what students need to learn) to assessment, because effective learning and

Figure 3.3 Next-generation learning system.
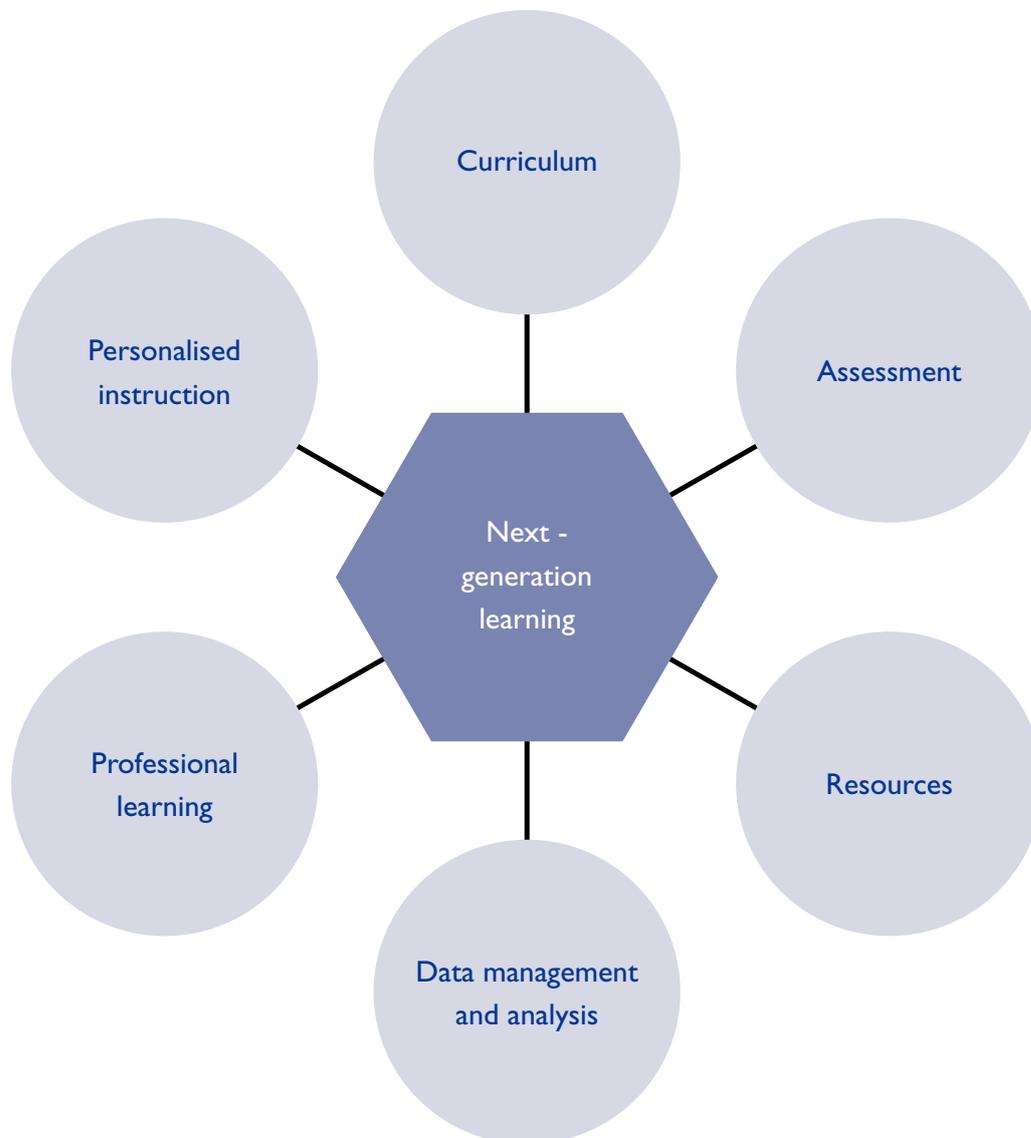
teaching require that one begins with the individual student and their starting points.

Geoff Masters quotes David Ausubel, the American psychologist renowned for his ground-breaking research into the role of advance organisers in learning, as having declared: 'If I had to reduce all of educational psychology to just one principle, I would say this: The most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly' (Ausubel 1968: vi, quoted in Masters 2013: 10).

So the primary role of assessment is to work out whether the student is ready to learn

the next segment of the curriculum, and, if not, where the gaps are so that these can be attended to first. As instruction proceeds, assessment is both backward-looking as a check on what has been learnt and on the quality of that learning and forward-looking in terms of readiness to tackle new content. Whereas in the past assessment has typically been looked upon as a discrete activity that follows teaching and learning, in the future it will be seen more as an aspect of ongoing instruction.

Assessment might take the form of a series of stand-alone mini-tests or quizzes, but, increasingly, it will be embedded naturally into learning activities so that assessment is continuous and unobtrusive, making use of the student's digital learning footprint to track progress, thereby encouraging immediate attention to learning obstacles if and when they are encountered and breaking down the barriers between learning and assessment.

Furthermore, such assessment will not always or even mainly be about assigning scores. As Sadler, one of the first to articulate the concept of formative assessment, observed many years ago: 'Qualitative [personalised] judgments are invariably involved in appraising a student's performance ... Growth takes place on many interrelated fronts at once and is continuous rather than lock-step' (1989: 123).

Through the use of rubrics, which will define performance in terms of a hierarchically ordered set of levels representing increasing quality of responses to specific tasks, and a common set of curriculum identifiers, it will be possible to not only provide immediate feedback to guide learning and teaching but also to build a digital record of achievement that can be interrogated for patterns and used to generate individualised and pictorial achievement maps or profiles.

Within next-generation learning systems, assessment will occur at all scales, from the most granular to the most synoptic. While its primary function will be formative, directed at proximal learning objectives and concerned with immediate feedback to improve learning and teaching, there will be a seamless transition to summative assessment of progress towards, and achievement of, wider curricular goals. What is more, these summative assessments will be demonstrably reliable, comparable and valid for incorporation into reporting systems, which can then support a range of uses including certification, selection and accountability. In other words, we see a new generation of assessments that will blur current distinctions and unhelpful dichotomies such as internal/external, formative/summative and qualitative/quantitative.

Much of the routine work in collecting, marking and extracting information from student responses will be automated, thus freeing up the teacher to focus on making use of the feedback obtained from daily observations and assessment tasks to personalise learning and improve instruction. An example of the kind of tool that makes this possible is Assistments, developed at the Worcester Polytechnic Institute.[11] Where professional judgement is involved in assessing work, multiple graders may be involved to ensure consistency of standards and to maximise the reliability of assessments.

While learning systems will embed a comprehensive range of assessments, authoring tools will also enable teachers to

---

11. See www.assistments.org (accessed 15 November 2014).

generate their own and upload them into the system for review and analysis as part of an overall development and quality-assurance process.

The assessment systems developed by PARCC and Smarter Balanced represent a significant milestone in the creation of large-scale, integrated online learning-assessment systems that incorporate assessments and tools to support formative classroom assessment practices, monitor student progress and meet mandatory accountability measures.

A further example of a more developmental initiative is New Pedagogies for Deep Learning (NPDL), a global partnership of clusters of 100 schools in each of ten countries that are committed to mobilising deep learning across systems.[12] One component of NPDL is a research-and-development effort to create a new generation of instruments and protocols to assess deep learning. The starting point will be setting out of competencies for learning tasks and assessing student progress. This will begin with adaptation of rubrics from the ITL Research/21CLD programme that defined levels and broad indicators of various deep-learning competencies.[13]

Resources

In generating instructional sequences, learning tasks and associated assessment activities, next-generation learning systems will embed or search out the resources that most closely match students' learning needs, accessing both purpose-built, commercially available materials and the rapidly expanding collections of public-domain/creative-commons resources.

The days of hard-copy textbooks, textbook-adoption regimes and the domination of the multibillion-dollar textbook market by a handful of publishers may be numbered. Many textbooks have been converted into digital format and made more interactive, thus bringing down costs, allowing more frequent updating of their contents and also opening up the field to smaller players.[14]

A plethora of interactive online resources is emerging, developed both commercially and by the profession itself. Much of this is being made available at low cost or free of charge. Examples of providers include KQED, a San Francisco-based public media outlet offering educators free resources for integrating media and new-media tools into teaching and learning, and CK12, a not-for-profit foundation that creates and aggregates high-quality resources aligned to state curriculum standards and offers its 'FlexBook System', an online platform for assembling, authoring and distributing interactive, multi-modal content for schools.[15]

Through meta-tagging of resources to the curriculum (facilitated by common terms and definitions) and also to other pertinent dimensions relevant to teaching, next-generation learning systems will tap into this much richer pool. For example, in Australia, education ministers have established Education Services Australia (ESA) as a not-for-profit company to support national priorities and initiatives and, in particular, to create, publish, disseminate and market curriculum and assessment materials, ICT-based solutions, products and services that support learning in the context

---

[12.] See http://www.newpedagogies.info (accessed 15 November 2014).

[13.] See http://www.itlresearch.com/itl-leap21 (accessed 15 November 2014).

[14.] See for example Boundless, with its online interactive textbook alternative that makes use of open-source content (www.boundless.com) and edSurge (www.edsurge.com/products/curriculum-products).

[15.] http://blogs.kqed.org/education/ and http://www.ck12.org/about/ (accessed 15 November 2014).

of a newly developed national curriculum. Through their 'Scootle' portal, ESA has created a 'one-stop shop' that provides teachers with access to more than 20,000 digital curriculum resources.[16] The content is indexed using an agreed vocabulary of curriculum topics and terms. Teachers can browse the Australian Curriculum and access appropriate, quality-assured digital resources that include activities for students, teacher support materials and interactive assessment resources.

Moreover, these resources look nothing like a traditional textbook. In an online world, they can take the form of interactive multimedia learning activities, games, videos, simulations, news articles, documentaries and so on. Or they may be short, simple ideas addressing a single, specific teaching/learning challenge, shared by practitioners in the field. Remotely located teachers and students engaged in learning and teaching the same or similar content can become a part of the total pool of resources that can be drawn upon to facilitate learning.

Next-generation resources require new and different quality-assurance processes. We must avoid teachers being lost in a sea of potentially useful resources without the capacity to locate and evaluate those most appropriate for the moment. So next-generation learning systems will incorporate ways to immediately locate quality resources directly relevant to specific aspects of the curriculum and the specific learning needs of a given group of students, as well as information on the efficiency and effectiveness of the resource in a given context.

Data management and analysis

It was not so long ago that almost all information about students and their learning was contained within teachers' books of marks, attendance registers, student record cards and student reports. Information on what was taught was in teachers' lesson plans, where these were available. But with the advent of computers in schools, most of this information has been systematised and digitised, and the amount of information collected has somewhat increased.

Next-generation learning systems, however, will create an explosion in data because they track learning and teaching at the individual student and lesson level every day in order to personalise and thus optimise learning. In an online world with intelligent software and a range of devices that facilitate unobtrusive classroom data collection in real time, the big challenges will lie not so much in obtaining data but in managing it and protecting privacy while turning it into powerful knowledge, something that data warehouses built just a few years ago were never designed to support.

Kristen DiCerbo and John Behrens (2014: 10) see these changes as amounting to a paradigm shift in assessment, involving:

- a focus on a broad range of attributes versus measuring narrowly defined knowledge and skills;
- integration of data over activity and time as opposed to over singular events;
- detailed tracking of context outside testing situations;
- dissolution of current distinctions such as 'informal' vs. 'formal' learning; and
- collection and permanence of learner profile data to make ongoing, intelligent recommendations.

[16.] http://www.esa.edu.au/projects/scootle (accessed 15 November 2014).

Next-generation learning systems will incorporate algorithms that interrogate assessment data on an ongoing basis and provide instant and detailed feedback into the learning and teaching process. Moreover, the information generated by learning systems will have value well beyond the individual learner: it will provide a source of generalisable new knowledge, paving the way for a 'design science' approach, in which the primary focus of educational research is on evidence-based strategies for improving learning and teaching.[17]

This will become increasingly viable through the application of data-mining and data analytics to discover patterns and relationships within the vast number of transactions that occur on a daily basis within classrooms.[18]

For so long, much of what happened inside classrooms has remained hidden in a 'black box', making it difficult to pursue a deliberate and continuous approach to the improvement of learning and teaching. Next-generation learning systems offer the prospect of revolutionising learning research and development by incorporating internal data-driven processes for improvement and by creating a design-focused concept of the role of research in shaping practice. In other words, we will see the development of learning systems consciously created as evolving products of ongoing research and development, aimed at achieving continuous improvement.[19]

Professional learning

In next-generation learning systems, the teacher retains the key role in fostering the learning for each student, but the job itself changes. Learning systems of the future will free up teacher time currently spent on preparation, marking and record-keeping and allow a greater focus on the professional roles of diagnosis, personalised instruction, scaffolding deep learning, motivation, guidance and care. This is the combination of activities that John Hattie describes as 'teacher as activator' (2009: 17).

Teachers will need to constantly update and acquire knowledge in order to perform this role effectively. They will need the kind of specific knowledge base characteristic of any true profession. Next-generation learning systems will therefore build in both formal and informal personalised professional learning for teachers, connecting them to instructional materials, resources and networks that provide timely, point-of-need professional development and support directly related to the task in hand, together with opportunities to gain recognition and credit for their learning and development.

Personalised instruction

With all the above in place, it is then possible to talk confidently about personalised instruction, which is the final and most crucial component of Figure 3.3. By personalised instruction, we mean instruction that is adjusted on a daily basis to the readiness of each student and that adapts to their specific learning needs, interests and aspirations. The fundamental premises of personalised learning have been a part of the writings of educators for decades but have become a realisable dream in recent years, thanks to the advent of new digital technologies.

[17.] One of the first and most persuasive to advocate a shift of the whole educational research enterprise towards improvement by design was Thomas Sergiovanni (see Sergiovanni 2000).

[18.] For a summary of this emerging field, see US Department of Education (2012).

[19.] This was foreseen by a number of writers a decade or more ago, notably by Professor David Cohen and colleagues (2003).

What does personalised instruction look like in practice? First and foremost, it means putting the individual student at the centre of the learning process and expecting them to achieve high standards. Second, it means better knowledge of learners, including not only detailed information on the specifics of what they already know but also about more generalised competencies, aptitudes, interests, aspirations and motivations. Third, it means learning goals that are specific to, and developed with, the awareness and involvement of the learner. Fourth, it means giving learners greater discretion in the learning activities and resources in which they will engage and adjusting teaching strategies to individual learners. Finally, it means expecting learners to take greater responsibility for their learning, to be more aware of their own strengths and weaknesses and to become actively engaged in the learning process.

Next-generation learning systems will assist the teacher in bringing together all the components needed to generate personalised instruction, including planning tools and a rich array of designed instructional materials, all specifically connected to relevant curriculum learning goals.

But the role of the teacher will have changed dramatically and may have become more differentiated. At its apex will be a new class of highly educated and trained professionals, expert in delivering personalised learning, with deep content and pedagogical knowledge, an intimate knowledge of each student, and knowledge and understanding of learning in a digital world.

RETHINKING, ALIGNING AND REBALANCING ASSESSMENT

This chapter has sought to provide a brief summary of some of the ways in which new thinking and digital technologies are transforming assessment and overcoming current barriers and limitations. Table 3.1 summarises what we see as the main features of this transformation.

In the case of formal assessment programmes created for the purposes of certifying student achievement, or for accountability purposes, these changes offer the prospect of significantly addressing some current limitations as identified in the previous chapter, providing assessments that are more able to:

- accommodate the full range of student abilities;
- provide meaningful information on learning outcomes;
- accommodate the full range of valued outcomes; and
- motivate improvement efforts and minimise opportunities for cheating and 'gaming' the system.

For assessment carried out as part of the ongoing process of learning and teaching, these changes offer exciting prospects too:

- a new generation of classroom-based learning and assessment activities capable of reliably assessing a much wider range of outcomes and generating instant and powerful feedback;
- assessment as an integral and vital part of sophisticated next-generation learning systems that will enable a new generation of empowered teachers to deliver personalised learning.

Table 3.1 Transforming assessment.

The ideal	How new thinking and technologies can help
Assessments that can accommodate the full range of student abilities	• Use of adaptive testing to generate more accurate estimates of student abilities across the full range of achievement while reducing testing time

Assessments that provide meaningful information on learning outcomes	• Online environments to facilitate:
	  – the administration of multiple versions of the same test in order to obtain information on performance across a much wider range of the curriculum
	  – the collection and analysis in real time of a wide range of information on multiple aspects of behaviour and proficiency and
	  – more immediate, detailed and meaningful reporting to specific stakeholder groups, such as via smartphone/tablet devices and through the creation of e-portfolios
	• Advances in the application of data analytics and the adoption of new metrics to generate deeper insights into and richer information on learning and teaching

Assessments that accommodate the full range of valued outcomes	• Automated marking to overcome obstacles to the more widespread use of essay and other open-response format questions • Platforms to support the delivery of a new generation of assessments specifically designed to assess deep learning and a range of inter- and intra-personal competences and character traits
Assessments that have integrity and are used in ways that motivate improvement efforts and that minimise opportunities for cheating and 'gaming' the system	• The adoption of (1) more cumulative approaches to approaches to assessment for selection purposes, with opportunities to re-sit; and (2) intelligent accountability systems that utilise multiple indicators of performance, that are designed to incentivise improvement and that avoid the creation of win–lose consequences for stakeholders for outcomes not fully under their control
Assessments that support students and teachers in making use of ongoing feedback to personalise instruction and improve learning and teaching	• Sophisticated online intelligent learning systems to integrate the key components involved in effective instruction and to support a new generation of empowered teachers in reliably assessing a much wider range of outcomes, using instant and powerful feedback on learning and teaching to deliver truly personalised instruction

That's quite an impressive list, but does it add up to an assessment renaissance? We believe that it does, but only if we are prepared to rethink some of the purposes of assessment, to seek a better alignment between assessment with curriculum and teaching and to rebalance assessment priorities.

An integrated, multi-level view of assessment

Perhaps the most urgent need right now in the field of assessment is for an overall conceptual framework and longer-term vision for its place and purpose in relation to the core processes of curriculum and of learning and teaching. We believe that the starting point is to think of assessment in an integrated, multi-level way, which, building upon the work of Rick Stiggins and Dale Duke (2008), and drawing upon earlier work by Peter Hill (2010), we represent as a three-level pyramid (see Figure 3.4).

Rather than focusing on discrete assessment programmes, we would suggest that it is more productive to view assessment as serving distinct data needs at three levels:

1 the teacher–student interface (traditionally the classroom);
2 the school; and
3 the system.

The most important level is the teacher–student interface, because this is where learning takes place and where there is the greatest need for assessment data to enable a truly personalised approach to learning and teaching. We would argue that the other two levels should be built on the assessment carried out at this first level.

Next is the school level, where education is managed and delivered. Schools need to draw

Figure 3.4 Tri-level assessment model.
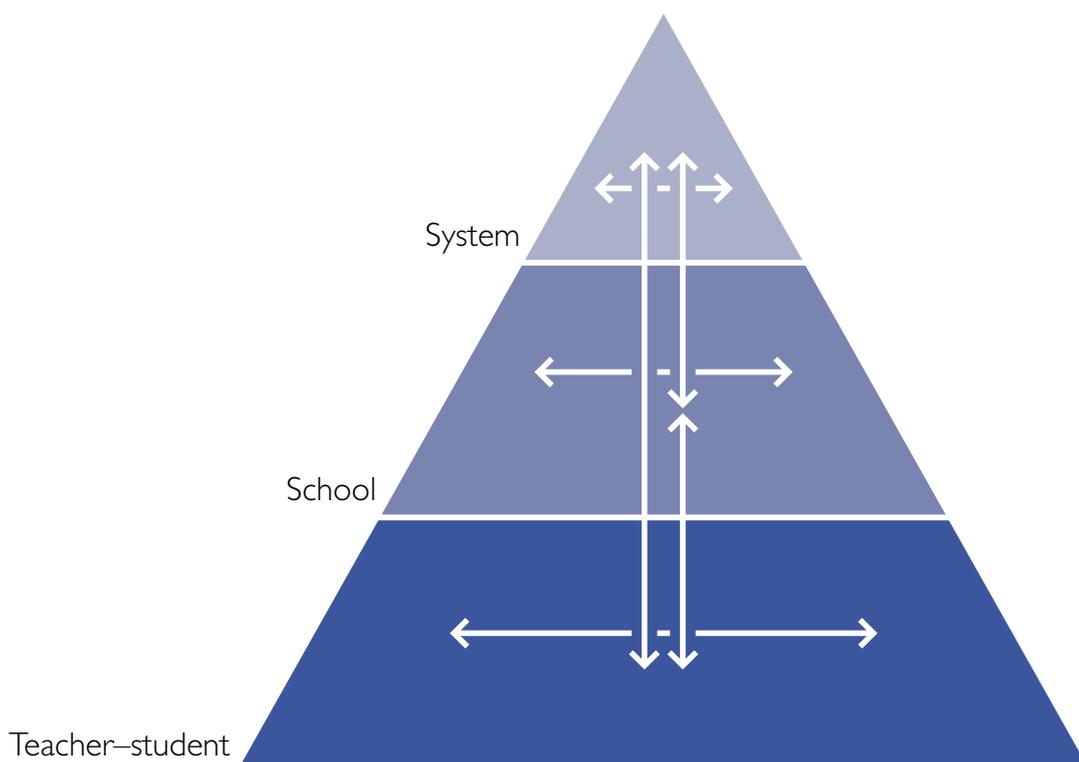
upon assessment data, collected at all three levels, to evaluate their performance, to be accountable to parents for the progress of their students and to manage learning and teaching within the school. This involves using assessment for both summative and formative purposes in addressing key questions such as:

- 'How are we doing relative to other schools?'
- 'Are we improving?'
- 'How successful are we in teaching the intended curriculum?'
- 'Which students, classrooms and teachers need extra support?'

At the top of the pyramid is the system that provides the policy and resourcing context for the schools it serves. Systems need assessment data for macro-level formative and summative purposes, including the evaluation of policies and programmes, to identify priorities and support needs, certifying student achievement, holding others to account and, in turn, being accountable for the performance of the system as a whole.

Within this tri-level assessment model, we envisage much greater vertical and horizontal flows of information among and within the three levels than currently occurs. We also predict greater reliance by systems on assessment carried out at the lower levels, as the availability and quality of assessment data collected at the teacher–student interface improves.

New developments in assessment, online assessment environments and next-generation learning systems provide the opportunity to rebalance assessment policies and practices so that they build on high-quality assessment of student progress at the teacher–learner interface, are fully aligned with the curriculum and with pedagogies adapted to twenty-first-century learning and support new and more sophisticated forms of certification and multi-level accountability. It requires close attention to the design not just of discrete assessments but to what Masters refers to as 'learning assessment systems' (2013: 32–56).

The challenge for awarding bodies

In considering the future of assessment for certification purposes, the challenge facing awarding bodies is to work out how they can take greater advantage of new technologies to deliver examinations online and thus improve their capacity to:

- assess a wider range of valued outcomes;
- create more authentic assessment tasks;
- assess the full range of student abilities more accurately and speed up the marking process, particularly for extended response questions;
- extend the window of time in which examinations may be taken and work towards the longer-term goal of examinations on demand; and
- use the potential of online assessment and developments in psychometric methods to more rigorously maintain and constantly benchmark standards to ensure they are world-class.

To date, many awarding bodies, while embracing onscreen marking, have moved only cautiously towards the adoption of online assessment, primarily due to constraints of connectivity and hardware availability. As these constraints are removed and solutions found to security and integrity issues, schools and students will increasingly opt for credentials offered via online assessment, noting that these credentials are less geographically

constrained than paper-based examinations and can be accessed anywhere in the world. This in turn may mark the beginning of the end for awarding bodies unable to invest in the infrastructure necessary to deliver world-class qualifications at low cost in an online environment. Clearly, the experience of providers of massive open online courses has relevance to awarding bodies at the senior secondary level and raises questions regarding the feasibility and desirability of high-quality international online credentials.

In addition, awarding bodies that serve geographically confined local jurisdictions need to consider the implications of globalisation, their ability to compete in the emerging global qualifications marketplace and whether they need to partner with other bodies in seeking to achieve best practice.

Certification needs to be conceptualised in ways that acknowledge the imperative for all students to be competent, continuous learners with the flexibility to respond to new life, work and study options and adapt successfully to rapid social, economic and technological change. Continuous learning clearly requires more dynamic approaches to certification and a greater willingness to assess and report the development of more generic competencies and relevant life experiences alongside the certification of formal learning.

The accountability challenges

In considering assessment for accountability purposes, the challenge for systems is to avoid or redress the mistake of implementing accountability systems that have high-stakes consequences for individuals, with decisions based primarily on results of short, poor-quality tests that assess a relatively narrow segment of the curriculum. Such systems typically create

perverse incentives, divert attention to the trivial and away from serious objectives and other more instructionally relevant uses of assessment, accelerate consumer distrust and fail to deliver expected improvements.

Getting the balance right is a key challenge in many parts of the world right now, although what this means in detail will vary significantly from country to country. It is worth quoting again, but more extensively this time, from the Gordon Commission (2013: 7–8) with reference to the US context, because it has relevance to accountability testing in many other countries:

The Commission calls on policymakers at all levels to actively promote this badly needed transformation in current assessment practice. The first and most important step in the right direction will require a fundamental shift in thinking about the purposes of assessment. Throughout the long history of educational assessment in the United States, it [assessment] has been seen by policymakers as a means of enforcing accountability for the performance of teachers and schools … But, as long as that remains their primary purpose, assessments will never fully realise their potential to guide and inform teaching and learning. Accountability is not the problem. The problem is that other purposes of assessment, such as providing instructionally relevant feedback to teachers and students, get lost when the sole goal of states is to use them to obtain an estimate of how much students have learned in the course of a year. It is critical that the nation's leaders recognise that there

are multiple purposes of assessment and that a better balance must be struck among them. The country must invest in the development of new types of assessments that work together in synergistic ways to effectively accomplish these different purposes – in essence, systems of assessment. Those systems must include tools that provide teachers with actionable information about their students and their practice in real time. We must also assure that, in serving accountability purposes, assessments external to the classroom will be designed and used to support high-quality education.

In other words, balance and alignment are critical when it comes to uses of assessment. The answer is not to abandon the search for rigorous systems of accountability but rather to engage the teaching profession in the design and implementation of systems that deserve their support.

An important avenue for building the profession's trust in accountability systems is through embracing the concept of reciprocal accountability, which Elmore states as implying that, 'For each unit of performance I demand of you, I have equal and reciprocal responsibility to provide you with a unit of capacity to produce that performance, if you do not already have that capacity' (2004: 244–5). The implications of reciprocal accountability for how systems and schools operate are substantial. Accountability is best thought of as a multi-level, shared, reciprocal process that all parties embrace.

Designing an effective accountability system involves clarifying who can and should be held

to account for what at each level of the system and establishing accountability arrangements that are reasonable, effective and promote a shared trust in the system. This means being sure, as far as possible, that accountabilities are within the power of the person or organisation being held to account.

In the school educational context, this typically means holding systems, schools and teachers responsible for:

- student growth or progress, rather than absolute levels of performance;[20] and
- doing those things that the evidence shows lead to improved outcomes – not just for achievement of the outcomes themselves (which may be only partly attributable to the specific person or organisation being held to account).

Direct accountability for outcomes is only appropriate where it is possible to separate out the impact of those being held to account. Having achieved agreement on accountability at different levels, one can then begin to align it with a multi-level system of assessment that balances out and aligns the claims of different purposes of assessment.

Equally important in the design of accountability systems is the need to take into account capacity-building requirements, particularly those related to teachers' assessment literacy and their capacity to make full use of the potential of assessment data, so that they can in turn provide feedback and enhance their own capacity to deliver more effective and personalised forms of learning and teaching.

The challenge for learning and teaching
This takes us to the challenges inherent in seeking to transform assessment undertaken

---

[20.] See, in particular, Betebenner and Linn (2010).

as part of the ongoing process of learning and teaching. Earlier, we noted the prospect of addressing the limitations of the age–grade progression model and of realising the potential of formative assessment in generating powerful feedback to optimise learning and teaching on a day-to-day basis. We suggested that this transformation would increasingly mean that formative assessment is an integral and vital part of learning systems designed to deliver personalised learning. We also proposed that this kind of assessment should provide the primary building block for all other kinds of assessment.

Such a transformation, we believe, is vital in order to break through the performance ceiling, significantly improve outcomes and reduce achievement gaps. However, it demands a huge change in thinking, upending more than a century of practice. Furthermore, the learning systems and technology required to support this kind of assessment are still in early development, so the transformation cannot be immediate. Nevertheless, it is already edging into a multitude of classrooms, typically as the result of the conviction and capacity of individual teachers, but sometimes with strong school or system support and direction. There is a growing consensus about the desirability of rejecting one-size-fits-all in favour of a personalised approach to learning, so long as it doesn't require extra resources and is feasible in typical classrooms. But there is considerable uncertainty about what the next steps to reaching that goal might be.

Becoming deeply involved in classroom assessment presents a challenge for systems that have not considered such assessment as a policy matter. It raises questions about the kinds of research and development needed to underpin quality assessment at this level and the systems required to collect and analyse the data such assessment provides. It also raises big issues about teacher development and teacher capacity in order to operate in a digital classroom in which the goal is personalised learning, with increasing integration of classroom activity into learning systems, and in which the teacher's role changes significantly – potentially in the direction of becoming more professional.

> ' we are on the verge of a radical change in thinking and practice regarding assessment in school education; … the exact form these changes will take depends very much on how we anticipate, plan for and shape them '

How does one prepare for such a future? As noted at the beginning of this essay, we are on the verge of a radical change in thinking and practice regarding assessment in school education. However, the exact form these changes will take depends very much on how we anticipate, plan for and shape them. This is the question that we address in the final section.

4. A FRAMEWORK FOR ACTION

In this chapter, we propose a way for policy-makers, schools, school-system leaders and other key players to prepare for the assessment renaissance, to ensure that they maximise the benefits of new developments and changes in thinking whilst avoiding the potential downsides. We present a framework for action that allows change to be implemented in ways and timeframes suited to the starting points, capacity and readiness of schools and systems. In the previous chapter, we focused on the potential benefits of the impending assessment renaissance, but it cannot be assumed that these benefits will always be realised. The path ahead is likely to be rocky. There are many examples of systems and schools that have had their fingers burnt by the over-hasty adoption of early and untried versions of next-generation assessment that failed to live up to expectations.

There are also examples of systems that have used assessment reform in ways that reinforce problematic practices and work against the more important, longer-term goals of personalising learning, enhancing teacher engagement and professionalism, incentivising students, teachers and school administrators and better aligning assessment with curriculum, learning and teaching.

As we indicated at the outset, while we may well be on the verge of a radical change in thinking and practice regarding assessment in school education, the exact form these changes will take depends very much on how we anticipate, envision, plan for and shape

them. Poorly executed, we could run into difficulties that take years to rectify.

In addition, we need always to be conscious of the wider context and of the fundamental changes that are happening in education more broadly, of which assessment is but one, albeit vital, part. That wider context will influence both the nature and the pace of change.

As we have indicated throughout, much of the innovation in the area of assessment will occur at the fringes of the system and perhaps outside it altogether, in the realm currently thought of as computer gaming. In addition, ideas and innovations will be shared laterally between schools and indeed across national boundaries. This process of innovation is to be welcomed, but an inevitable consequence, without intervention, would be haphazard adoption and potentially a growing gap between the 'haves' of the assessment re-naissance and the 'have nots'. If there is to be universal benefit across a system, governments will need to act. Moreover, some of what is required to make the renaissance universal, such as the technological infrastructure, cannot be provided by individual schools.

The realisation of the assessment renaissance and its benefits depends, therefore, on governments, systems and schools playing a powerful strategic role. Here we set out what the key features of that role might be.

1. THINK LONG-TERM

The assessment renaissance, we firmly believe, is coming. But it is hard to predict when it will arrive. Among parents and other stakeholders there are strong attachments to the status quo, and the technical challenges of universal implementation remain formidable.

In these circumstances, it is essential to think long-term. Since it is also hard to predict precisely how the assessment renaissance will occur, governments must keep their options open, while at the same time investing in the capacity to bring about the assessment renaissance – that is the research, the experimentation, teachers' skills and the technology – and maintaining close connections both to what is happening in the field and to what is happening internationally. System leaders need to encourage assessment developers and awarding bodies to experiment. At some point it may also be necessary to lift some current regulations in order to enable the kind of experimentation required. Ministers and top officials might also begin to explore in public how, say, over a ten-year period, assessment might develop. While there are always risks in doing so – it is especially important not to devalue existing qualifications – holding out a vision of a transformed system is also important.

2. BUILD PARTNERSHIPS

Bringing about the assessment revolution requires collaboration, certainly between the teaching profession and government but also between other key players such as education and technology companies, edtech venture capitalists and university researchers. Schools need to seek to collaborate with and learn from each other and to promote change through participation in local, national or even international school networks.

To some extent, the necessary collaboration can be expected to occur organically, but it could be accelerated and deepened if it is incentivised. For example, in a report for the Massachusetts Business Alliance for Education in which we were both involved (MBAE 2014), it was recommended that the state run an annual competition, the Massachusetts Accelerated Learning Challenge, through which educators, technology innovators and venture capitalists would be incentivised to propose solutions to the state's priority challenges, which might include radical innovation in assessment. Well-run competitions have the benefit of not just developing the solutions that come from the winners but also of creating the relationships among key players who don't win but who might go on to collaborate. Similar benefits apply at the school level, where competitions can incentivise staff to explore new kinds of pedagogies and assessment in teaching the curriculum and thus encourage and awaken interest and awareness among colleagues.

As Michael Fullan and Katelyn Donnelly argue in *Alive in the Swamp* (2013), digital innovations in school systems are likely to require simultaneous action in relation to three elements – system change, pedagogy and the technology itself. Governments, systems and school leaders need to ensure that they have grasped this key point conceptually and that they encourage the collaboration that will enable all three angles to be worked on simultaneously. Often this will involve building consortia that pool expertise.

3. CREATE THE INFRASTRUCTURE

One barrier to progress on assessment and other digital transformations is the poor quality of the technological infrastructure, which affects connectivity and the high reliability required for effective online assessment. At each level in the system, including at individual school level, there needs to be a chief information officer who deeply understands both education and technology and who can ensure the necessary infrastructure, hardware and software. It is not always necessary for an individual system or school to work all this out for itself – sharing of expertise and benefits of scale suggest, again, collaboration and consortia. But the infrastructure, hardware, software and maintenance are all critical.

4. DEVELOP TEACHER CAPACITY

In the *Alive in the Swamp* triangle, often the teacher capacity to change pedagogy lags behind the digital infrastructure. We have seen earlier in this paper how teacher capacity to deploy assessment for learning remains constrained by circumstances and, in some cases, a lack of the professional skills required. Any sensibly long-term strategy would invest, over, say, a five-year period, in developing teachers' familiarity with both the technology and sophisticated assessment. These skills should largely be developed at school level through coaching and mentoring, sometimes using sequences of video to demonstrate approaches and evaluate skills. This, is turn, means that the starting point for any system-wide approach should be the development of the necessary awareness among principals, who will often be part of multi-school networks.

5. ALLOW VARIATION IN IMPLEMENTATION

With some educational change, it is necessary for the whole system or school to move in lockstep. With the assessment renaissance, we recommend a different approach: allow variation, encourage schools, networks of schools and individual teachers to innovate with a framework and learn from the most successful examples. At critical points, the whole system or school may need to move in unison, but in most of the world that moment has not yet arrived.

To be clear, though, we are not recommending simply leaving the system, school or teacher alone and seeing what happens. On the contrary, we are suggesting a strategic approach, overseen by government, working within a framework and designed to learn quickly and effectively from a variety of approaches.

6. ADOPT A DELIVERY APPROACH

The potential of the assessment renaissance and the need for sustained implementation over a decade mean that it makes sense to apply a 'delivery' approach: make it a priority, plan ahead, ensure routine check-ins with all key players and make clear who is responsible.[1] Solve problems as they arise. Do so conscious of what is happening elsewhere in the world, and ensure systematic learning from it.

In many countries, over such a sustained period, there will be changes of government following elections and, even more likely, changes of minister. It is always a setback if assessment becomes politically polarised, because if the approach keeps changing, the benefits of any assessment, however good, are undermined

---

[1.] See Barber with Moffit and Kihn (2011).

by the uncertainty. It is particularly damaging if qualifications, which need to be a currency in the labour market, become politically contested. For these reasons, as far as possible, governments should strive to gain cross-party consensus for assessment strategy and thus enable it to be pursued systematically over an extended period. Much the same issues apply at the level of system and school leaders, although generally with less force, meaning that it is important to create a shared vision owned by all rather than by the current leadership.

7. COMMUNICATE CONSISTENTLY

Earlier we identified the 'forty-year communication gap'. There are many misconceptions in the assessment debate, especially (but not only) among parents and the public. There is a strong attachment to traditional assessment in many countries, including China, Korea and the UK. Government and educators often add to the confusion by engaging in loud and sometimes wilfully misleading debate. If the assessment renaissance is to come about and its benefits for learners are to be realised, then there will need to be consistent communication, ideally with government and leading educators working together on the messages, and with school principals and teachers communicating to parents the significance of the changes for their children.

8. APPLY THE CHANGE KNOWLEDGE

In approaching the task of change management, our starting point needs to be our knowledge base of what it takes to achieve successful, system-wide change. We summarise this knowledge base below, adapted from a set of conclusions that we previously published

(Barber and Fullan 2005) and supplemented with additional conclusions of specific relevance to assessment reform. We call these conclusions a 'Tri-Level Reform Solution', because we consider them relevant to the aforementioned three levels of teacher–learner, school and system.

Moral purpose
The overwhelming majority of educators are motivated by a sense of moral purpose. This applies particularly to the role of assessment. Moral purpose is heightened when assessment is seen as the key to improving learning, especially for those who are falling behind, or to providing recognition of student achievement.

Positive experiences
People frequently change their behaviours before they change their beliefs. New, positive experiences with next-generation assessment will be a powerful motivator, especially when they relate to fulfilling moral purpose. Moreover, they will differ from individual to individual, depending on their starting point.

Shared vision and ownership
Motivation is further enhanced when there is a shared vision and ownership of change. Successful systems and schools don't simply demand change; they build a shared vision and ownership and engage all stakeholders in its creation and realisation. Next-generation assessment must be willingly embraced by the profession rather than imposed from above.

Learning in context is key
Even the best professional development workshops are only input for success. Actual success occurs in the context of daily learning. The most fundamental feature of next-generation assessment – its use to improve learning and teaching – can only be

understood by learning about it in the context of daily classroom learning and teaching.

Encourage and learn from the pioneers

Next-generation assessment is more a movement than a defined change. It will move forwards on many different fronts, and not all will ultimately prove fruitful. It is important to encourage and reward the pioneers and the risk-takers, and to learn from them.

Professional learning communities at the school and school-network levels are crucial in establishing purposeful and collaborative learning cultures in which teachers learn from each other and school leaders and teachers collaborate for continuous improvement. For next-generation assessment to become a reality, teachers will need to adopt, over time, a different and more professional role than the one currently demanded by one-size-fits-all instructional approaches. Professional learning communities are the key to bringing about this role transformation. In addition, as Michael Fullan would argue, professional learning in a purposeful and collaborative learning culture can be a powerful way to reduce ineffective teaching and unwanted variation and maximise effective teaching and positive variation.

System support

Schools, their leaders and the professional learning communities within them will not be sustained unless the system actively supports and encourages them and fosters and maintains their development. While some systems are still struggling with the infrastructure issues of interconnectivity and hardware, others are grappling with problems such as identifying open platforms for next-generation learning systems, accessing quality online content, designing new assessments and so on.

Balance pressure and support

Systems and schools must integrate pressure and support so that there is serious engagement in capacity-building with a focus on efficacy. Capacity-building is what many policy-makers and system and school leaders neglect, but it is vital when it comes to next-generation assessment, which is about enhancing capacity, not reducing the need for it.

Lateral capacity is vital for spreading knowledge and increasing commitment. Lateral capacity-building consists of strategies that enable teachers, schools and school systems to learn from each other. This implies systematic and purposeful networking to connect with those who are on the same journey, but perhaps in a different place on the path.

Leadership is the key to system transformation

Leaders must work with a vision, goals and more proximal objectives and do so with and through the development of other leaders as they go. It also means having leaders with specialist knowledge of the field, such as a full-time chief information officer whose role is to attend to digital needs and the use of technology to improve learning and teaching.

Better value for money

The logistical complexity and costs of most current formal assessment programmes are formidable. Apart from test development, they include:

- printing the tests;
- maintaining the security of printed tests;
- secure distribution and collection of papers;
- labour-intensive marking of scripts;
- data entry and cleaning;

- psychometric work to calibrate test items, equate tests and generate results;
- preparing results for publication and making them available to schools and a wider public along with relevant advice; and
- providing support materials to assist stakeholders in making use of the data.

Once schools and homes have connectivity and the relevant hardware to support online assessment, and once systems invest in more sophisticated test-delivery systems, the burden of a number of these logistical and cost issues can be reduced significantly.

Fullan and Langworthy (2014) provide a compelling argument that, while costs are coming down every day, even at current prices, the costs per student per year can be offset through reprioritisation and savings in other areas. In the case of assessment, there are specific additional upfront costs in developing relevant software, creating quality banks of items and creating new kinds of tests or examinations. However, considerable savings are possible through work with other systems that have already done or are about to do this developmental work and are prepared to share it at little or no cost.

But these costs need to be considered alongside the expected benefits and, in particular, the significantly higher learning outcomes achievable by using online assessment to facilitate formative assessment and generate instructionally valuable feedback. Professor John Hattie's meta-analysis of the research literature (2009) indicates their sizeable effect (sizes in excess of 0.7 of a standard deviation). In other words, the level of investment in online assessment and in building teacher capacity required to facilitate and realise the benefits

of formative assessment and feedback is small relative to the potential pay-off in learning outcomes.

DRAWING TOGETHER THE THREADS

Our argument has been that the 'push' factor of globalisation and the 'pull' factor of the performance ceiling are together giving rise to an educational revolution in which certain long-held beliefs and ways of doing things are repudiated and replaced by a new set of beliefs and practices.

The seeds of each of these key changes can be seen all around us, but full adoption will take some time to achieve. And for the education revolution to happen, we will have to change our views on the following factors:

- A student's capacity to learn and profit from formal education.
- What students need to learn. There has to be a greater emphasis on the deeper understanding of big ideas, the organising principles of disciplines and explicit and systematic attention to twenty-first-century skills.
- The focus of educational policy. We need a shift from focusing on the school to focusing on the individual student.
- The basic organisation of schooling, in particular a repudiation of the age–grade progression model in favour of access and progression more aligned to a student's readiness to learn.
- How students will learn and how teachers will teach, in particular, a shift towards much of learning time spent within an online learning environment, with teachers focused less on providing knowledge and more on assisting students to apply their knowledge,

enabling them to overcome barriers to progress and helping them to discern what is important and true.

- The emergence of teaching as a true profession, with a distinctive knowledge base, a framework with well-defined common terms for describing and analysing teaching and strict control by the profession itself on entry into and advancement within teaching.

'there is consensus among leaders in the field that we are on the brink of an assessment renaissance that will help secure high standards for all, remove current achievement ceilings and support a focus on the higher-order thinking and inter- and intra-personal skills vital for living and learning in the twenty-first century '

We have argued that when moving in these directions, assessment tends to be controversial, the lagging factor and a barrier to change. However, there is consensus among leaders in the field that we are on the brink of an assessment renaissance that will help secure high standards for all, remove current achievement ceilings and support a focus on the higher-order thinking and inter- and intra-personal skills vital for living and learning in the twenty-first century.

In the case of formal assessment programmes designed primarily for certification, selection and accountability purposes, there is the prospect of creating tests and examinations that:

- assess the full range of student abilities;
- provide more meaningful information on learning outcomes;
- assess the full range of valued outcomes;

- motivate improvement efforts; and
- minimise opportunities for cheating and 'gaming' the system.

In the case of assessment carried out as part of the ongoing process of learning and teaching, these changes bring the possibility of:

- a new generation of classroom-based learning and assessment activities capable of reliably assessing a much wider range of outcomes and generating instant and powerful feedback; and
- assessment that is integrated into sophisticated, next-generation learning systems that enable a new cadre of empowered teachers to deliver personalised learning.

Realising these benefits will not be easy. Moreover, it must be remembered that changes to assessment are taking place as part of even more fundamental changes in education. This wider context will affect both the nature and the pace of change.

With this context in mind, we advocate the adoption of the following framework for action.

- Think and plan for the longer-term.
- Build partnerships.
- Create the necessary infrastructure.
- Develop teacher capacity.
- Allow variation in implementation.
- Adopt a delivery approach.
- Apply the knowledge we already have about the process of change.

Above all, we believe it is vital not to underestimate the significance of what is taking place in this field. We see these changes in thinking on assessment leading to a veritable

renaissance, a revival in thinking and practice that promises to overcome many of the key limitations of the current paradigm and put assessment more fully in the service of the curriculum and of learning and teaching. And, for this to happen, governments, systems, schools and those within them all have critical roles to play.

REFERENCES

Ausubel, D. P. (1968) *Educational Psychology, A Cognitive View*, New York: Holt, Rinehart & Winston.

Barber, M. (2014) 'Consistent Quality Plus Innovation: Not One or the Other, Both', Education Reform Summit, 10 July. Available at http://blog.pearson.com/wp-content/uploads/2014/07/20140710-National-Education-Reform-Summit-2.pdf (accessed 12 November 2014).

Barber, M., K. Donnelly and S. Rizvi (2012) *Oceans of Innovation: The Atlantic, the Pacific, Global Leadership and the Future of Education*, London: IPPR. Available at http://www.ippr.org/publication/55/9543/oceans-of-innovation-the-atlantic-the-pacific-global-leadership-and-the-future-of-education (accessed 12 November 2014).

Barber, M. and M. Fullan (2005) 'Tri-level Development: It's the System', *Education Week*, 2 March.

Barber, M., with A. Moffit and P. Kihn (2011) *Deliverology 101: A Field Guide for Educational Leaders*, Thousand Oaks, Calif.: Corwin.

Barber, M. and M. Mourshed (2007) *How the World's Best-Performing School Systems Come Out on Top*, London: McKinsey & Company. Available at http://mckinseyonsociety.com/how-the-worlds-best-performing-schools-come-out-on-top (accessed 12 November 2014).

Betebenner, D. W. and R. L. Linn (2010) *Growth in Student Achievement: Issues of Measurement, Longitudinal Data Analysis and Accountability*, Princeton, NJ: K-12 Assessment and Performance Management Center, ETS.

Black, P. and D. Wiliam (1998) 'Assessment and Classroom Learning', *Assessment in Education*, 5 (1): 7–71.

Chadowsky, N. and V. Chadowsky (2010) *State Test Score Trends through 2008–09, Part 1: Rising Scores on State Tests and NAEP*, Washington, DC: Center on Education Policy. Available at http://www.cep-dc.org/publications/index.cfm?selectedYear=2010 (accessed 18 November 2014).

Clesham, R. (2013) *Good Assessment by Design: An International Comparative Analysis of Science and Mathematics Assessments*, London: Pearson. Available at http://uk.pearson.com/content/dam/ped/pei/uk/pearson-uk/Documents/wcq/good_assessment_by_design_report.pdf (accessed 12 November 2014).

Cohen, D. K., S. W. Raudenbush and D. L. Ball (2003) 'Resources, Instruction and Research', *Educational Evaluation and Policy Analysis*, 25 (2): 119–42.

Conley, D. T. and L. Darling-Hammond (2013) *Creating Systems of Assessment for Deeper Learning*, Palo Alto, Calif.: Stanford Center for Opportunity Policy in Education.

DiCerbo, K. E. and J. T. Behrens (2014) *Impacts of the Digital Ocean on Education*, London: Pearson. Available at https://research.pearson.com/content/plc/prkc/uk/open-ideas/en/articles/a-tidal-wave-of-data/_jcr_content/par/articledownloadcompo/file.res/3897.Digital_Ocean_web.pdf (accessed 12 November 2014).

Dikli, S. (2006) 'An Overview of Automated Essay Scoring', *Journal of Technology, Learning and Assessment*, 5 (1). Available at http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1640 (accessed 12 November 2014).

Dweck, C. S. (2006) *Mindset: The New Psychology of Success*, New York: Random House.

Elmore, R. F. (2004) *School Reform from the Inside Out: Policy, Practice and Performance*, Cambridge, Mass.: Harvard Education Press.

European Commission (2013) *Commission Staff Working Document: Digital Agenda Scoreboards 2013*, available at http://ec.europa.eu/digital-agenda/sites/digital-agenda/files/DAE%20SCOREBOARD%202013%20-%20SWD%202013%20217%20FINAL.pdf (accessed 15 November 2014).

Feith, D. (2011) *Teaching America: The Case for Civic Education*, Lanham, Md.: Rowman & Littlefield.

Fullan, M. (2008) *The Six Secrets of Change: What the Best Leaders Do to Help Their Organizations Survive and Thrive*, San Francisco, Calif.: Jossey-Bass.

Fullan, M. and K. Donnelly (2013) *Alive in the Swamp: Assessing Digital Innovations in Education*, London: Nesta. Available at http://www.nesta.org.uk/sites/default/files/alive_in_the_swamp.pdf (accessed 12 November 2014).

Fullan, M., P. Hill and C. Crévola (2006) *Breakthrough*, Thousand Oaks, Calif.: Corwin Press.

Fullan, M. and M. Langworthy (2014) *A Rich Seam: How New Pedagogies Find Deep Learning*, London: Pearson. Available at http://www.michaelfullan.ca/wp-content/uploads/2014/01/3897.Rich_Seam_web.pdf (accessed 12 November 2014).

Global Education Leaders Program (2014) 'Transforming Global Education with New Metrics: Statement by the Global Education Leaders' Program, available at http://gelponline.org/sites/default/files/resource-files/gelp_statement_june_2014.pdf (accessed 15 November 2014).

REFERENCES

Gordon Commission on the Future of Assessment in Education (2013) *A Public Policy Statement*, Princeton, NJ: The Gordon Commission. Available at http://www.gordoncommission.org/rsc/pdfs/gordon_commission_public_policy_report.pdf (accessed 12 November 2014).

Grossman P. and M. McDonald (2008) 'Back to the Future: Directions for Research in Teaching and Teacher Education', *American Educational Research Journal*, 45 (1): 184–205.

Hannon, V., A. Patton and J. Temperley (2011) *Developing an Innovation Ecosystem for Education*, San Jose, Calif.: Cisco Systems. Available at http://www.cisco.com/web/strategy/docs/education/ecosystem_for_edu.pdf (accessed 12 November 2014).

Hattie, J. (2009) *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*, London and New York: Routledge.

— (2012) *Visible Learning for Teachers: Maximizing Impact on Learning*, London and New York: Routledge.

Hattie, J. and H. Timperley (2007) 'The Power of Feedback', *Review of Educational Research*, 77 (1): 81–112.

Heylighten, F. (2012) 'Conceptions of a Global Brain: An Historical Review', in B. Rodrigue, L. Grinin and A. Korotayev (eds), *From Big Bang to Global Civilization: A Big History Anthology*, Berkeley, Calif.: University of California Press. Available at http://pcp.vub.ac.be/papers/GB-Conceptions-Rodrigue.pdf (accessed 13 November 2014).

Hill, P. W. (2010) 'Using Assessment Data to Lead Teaching and Learning', in A. M. Blankstein, P. D. Houston and R. W. Cole (eds.), *Data-Enhanced Leadership: Using What You Know to Be a More Effective Leader*, Thousand Oaks, Calif.: Corwin Press, pp. 31–50.

Hill, P. W. and K. J. Rowe (1996) 'Multilevel Modeling in School Effectiveness Research', *School Effectiveness and School Improvement*, 7 (1): 1–34.

Ho, A. D. (2008) 'The Problem with "Proficiency": Limitations of Statistics and Policy under No Child Left Behind', *Educational Researcher*, 37 (6): 351–60.

Hursh, D. (2007) 'Assessing No Child Left Behind and the Rise of Neoliberal Education', *American Educational Research Journal*, 44 (3): 493–518.

Leadbeater, C. (2002) *Learning about Personalization*, London: Innovation Unit, Department for Education and Skills.

Levitt, S. D. and S. J. Dubner (2007) *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*, London: Penguin Books.

Lissitz, R. W. and H. Jiao (eds.) (2012) *Computers and Their Impact on State Assessments*, Charlotte, NC: Information Age Publishing, Inc.

Lortie, D. C. (1975) *Schoolteacher: A Sociological Study*, Chicago, Ill.: University of Chicago Press.

Massachusetts Business Alliance for Education (2014) *The New Opportunity to Lead: A Vision for Education in Massachusetts in the Next 20 Years.* Available at http://www.mbae.org/wp-content/uploads/2014/03/New-Opportunity-To-Lead.pdf (accessed 13 November 2014).

Masters, G. N. (2013) *Reforming Educational Assessment: Imperatives, Principles and Challenges*, Camberwell, Victoria: Australian Council for Educational Research. Available at http://research.acer.edu.au/aer/12 (accessed 13 November 2014).

Mehta, J. (2013) *The Allure or Order: High Hopes and Dashed Expectations and the Troubled Quest to Remake American Schooling*, Oxford: Oxford University Press.

Mislevy, R. (2013) 'Postmodern Test Theory', in The Gordon Commission, *To Assess, To Teach, To Learn: A Vision for the Future of Assessment – Technical Report.* Available at http://www.gordoncommission.org/rsc/pdfs/gordon_commission_technical_report.pdf (accessed 13 November 2014).

Mislevy, R. J., J. T. Behrens, K. E. Dicerbo and R. Levy (2012) 'Design and Discovery in Educational Assessment – Evidence-Centered Design, Psychometrics, and Educational Data Mining', *Journal of Educational Data Mining*, 4 (1): Article 2. Available at http://researchnetwork.pearson.com/wp-content/uploads/mislevyetalvol4issue1p11_48.pdf (accessed 13 November 2014).

Newton, P. (2007) 'Clarifying the Purposes of Educational Assessment', *Assessment in Education*, 14 (2): 149–70.

Ng, P. T. (2008) 'Educational Reform in Singapore: From Quantity to Quality', *Educational Research for Policy and Practice*, 7: 5–15.

Oates, T. (2013) *Tiering in GCSE: Which Structure Holds Most Promise?* Available at http://www.cambridgeassessment.org.uk/Images/138921-tiering-in-gcse-which-structure-holds-most-promise-.pdf (accessed 13 November 2014).

OECD (2008) *21st Century Learning: Research, Innovation and Policy Directions from Recent OECD Analyses.* Available at http://www.oecd.org/dataoecd/39/8/40554299.pdf (accessed 13 November 2014).

— (2010) *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices (Volume IV).* Available at http://dx.doi.org/10.1787/9789264091559-en (accessed 13 November 2014).

REFERENCES

— (2011) *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices*, vol. IV. Available at http://dx.doi.org/10.1787/888932343285 (accessed 24 November 2014).

— (2012) *PISA 2009 Technical Report*, Paris: OECD Publishing. Available at http://dx.doi.org/10.1787/9789264167872-en (accessed 27 November 2014).

— (2013a) *Draft Collaborative Problem-Solving Framework*. Available at http://www.oecd.org/pisa/pisaproducts/pisa2015draftframeworks.htm (accessed 18 November 2014).

— (2013b) *Pisa 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science*, vol. I, Tables 1.4.3b and 1.2.3b, Annex B1. Available at http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-i.htm (accessed 13 November 2014).

— (2013c) *PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices*, vol. IV. Available at http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-iv.htm (accessed 18 November 2014).

Pellegrino, J. W., M. L. Hilton, Committee on Defining Deeper Learning and 21st Century Skills (2012) *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*, Washington, DC: National Research Council of the National Academies. Available at http://www.leg.state.vt.us/WorkGroups/EdOp/Education%20for%20Life%20and%20Work-%20National%20Academy%20of%20Sciences.pdf (accessed 18 November 2014).

Phillips, T. (2013) 'Bras Begone: China Clamps Down on Cheating in University Entrance Exams by Banning Brassieres', *The Telegraph*, available at http://news.nationalpost.com/2013/06/06/bras-begone-china-clamps-down-on-cheating-in-university-entrance-exams-by-banning-brassieres (accessed 15 November 2014).

Pierson, R. (2011) *Learning Ecosystem Brief Design Paper: Mass Customized, Personalized Learning*. Available at http://www.innovationunit.org/sites/default/files/Pierson paper.pdf (accessed 13 November 2014).

Polokoff, M. S., A. J. McEachin, S. L. Wrabel and M. Duque (2014) 'The Waive of the Future? School Accountability in the Waiver Era', *Educational Researcher,* 43 (1): 45–54. Available at http://edr.sagepub.com/content/43/1/45.full.pdf+html?ijkey=LoPEgefArEO0M&keytype=ref&siteid=spedr (accessed 13 November 2014).

Popham, W. J. (2008) *Transformative Assessment*, Alexandria, Va.: Association for Supervision and Curriculum Development.

Sadler, D. R. (1989) 'Formative Assessment in the Design of Instructional Systems', *Instructional Science*, 18: 119–44.

Sergiovanni, T. J. (2000) 'Changing Change: Towards a Design Science and Art', *Journal of Educational Change*, 1 (1): 57–75.

Soland, J., L. Hamilton and B. M. Stecher (2013) 'Measuring 21st-Century Skills: Guidance for Educators', Working Paper, Rand Education. Available at http://www.rand.org/pubs/external-publications/EP50463.html (accessed 13 November 2014).

Statistics Commission (2005) *Measuring Standards in English Primary Schools*, London: Statistics Commission.

Stiggins, R. and D. Duke (2008) 'Effective Instructional Leadership Requires Assessment Leadership', *Phi Delta Kappan*, 90 (4): 285–91.

Tucker, M. (2013a) 'Linn and Everson on Testing, Standards and Accountability', available at http://blogs.edweek.org/edweek/top_performers/2013/09/linn_and_everson_on_testing_standards_and_accountability.html (accessed 15 November 2014).

— (2013b) *Why Has US Education Performance Flatlined?* Available at http://blogs.edweek.org/edweek/top_performers/2013/12/why_has_us_education_performance_flatlined.html (accessed 13 November 2014).

US Department of Education (2012) 'Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief', available at http://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf (accessed 15 November 2014).

Wigdor, A. K. and B. F. Green Jr. (eds.) (1991) *Performance Assessment for the Workplace*, vol. l. Committee on the Performance of Military Personnel, Commission on Behavioral and Social Sciences and Education, National Research Council, Washington, DC: National Academy Press.

Willingham, D. T. (2006) 'How Knowledge Helps: It Speeds and Strengthens Reading Comprehension, Learning – and Thinking', *American Educator*. Available at http://www.aft.org/newspubs/periodicals/ae/spring2006/willingham.cfm (accessed 13 November 2014).

Zhang, M. (2013) 'Contrasting Automated and Human Scoring', *R&D Connections*, 21 (March). Available at http://www.ets.org/Media/Research/pdf/RD_Connections_21.pdf (accessed 13 November 2014).

ALWAYS LEARNING

PEARSON