Stimulus Paper

Foreword

In recent years, it has become evident that attitudes towards testing and assessment in schools have become quite negative. This has been true particularly in the USA, but also applies in England. Key stakeholders have begun to see certain kinds of assessment as counter productive and assessment providers as part of the problem rather than part of the solution.

As specialists in the fields of assessment and education, Pearson believes it has both a key contribution to make to address these concerns, and an obligation to do so. This paper shares some of our initial research and sets out thoughts to serve as the basis for a wider discussion with the education community.

As a first step, we commissioned a survey among a random sample of parents, teachers and head teachers. The survey confirmed that testing and assessment have become controversial and that those surveyed hold quite negative attitudes. For example, parents complain that the only thing teachers care about now is test preparation and test results, while teachers complain that high stakes assessment is distorting the curriculum and leads to pressures to game the system rather than improve teaching and learning.

Our initial analysis of these attitudes suggests that concerns go deeper than those that Pearson alone can address and involve fundamental questions regarding the purposes of assessment and policy regarding how assessment data are used.

This led to the realization that we need to go further and that the best way for us to stimulate thinking about these issues would be initiate an open consultation with key stakeholders to better understand concerns and identify a way forward to address them.

To this end, this paper seeks to generate informed discussion and debate about the future of assessment in England.[1] In particular, we seek your ideas, advice and reactions to proposals for specific ways in which Pearson can respond to the expressed concerns and perhaps facilitate debate and action more generally that will lead to improvements in educational assessment in this country.

Two important riders:

1. Chatham House rules apply. While we may publish outcomes of the consultation events, we will not either explicitly or implicitly attribute any words or ideas to any individual. Any quotations cited will be anonymous.
2. At Pearson we recognise that while we can facilitate discussion and lend our expertise in this field, it is the opinions of teachers that count the most. The objective of this exercise is to gather these opinions, engage with parents and

[1] The paper was prepared by Dr Peter Hill and draws freely from Hill, P. & Barber, M. (2014). *Preparing for a renaissance in assessment*. London: Pearson.

put forward constructive new thinking to policy makers.

Finally, I would like to thank Dr Peter Hill who has prepared this stimulus paper, and all of you who have kindly agreed to contribute your thoughts, experience and expertise.

Rod Bristow,
President, Pearson UK and Core Markets

Overview

In acknowledgement of negative attitudes to testing and assessment, particularly those expressed in an initial survey by a sample of parents, teachers and head teachers, a need was identified within Pearson for a more in-depth open consultation with key stakeholders to better understand concerns and identify ways that might address them. This paper seeks to stimulate discussion in the context of these consultations and building on the results of the initial survey.

The paper poses 15 discussion points that address:

- the extent to which stakeholders share similar views to those expressed in the initial survey;
- whether negative attitudes to accountability testing are inescapable or could be ameliorated through a better balance in the emphasis placed on different or new kinds of assessment;
- attitudes to assessment for learning and its relationship to summative assessment;
- attitudes to assessment for accountability purposes;
- desirable characteristics of accountability systems;
- whether accountability testing practices in other countries provide pointers to how to best mitigate negative impacts of accountability testing; and
- what might be done over the longer term to seek to resolve tensions between validity and reliability in accountability testing.

No specific proposals are offered within the paper as its purpose is to better understand stakeholders' views and concerns and to stimulate discussion on possible ways of generating more positive attitudes towards testing and assessment in the longer term.

Attitudes towards assessment

In June 2015, *BritainThinks* conducted on behalf of Pearson a survey amongst a random sample of around 500 parents, teachers and head teachers[2] of attitudes to testing and assessment. Its purpose was to probe the following hypotheses:

- There are concerns among teachers and parents alike about testing and assessments in English schools.

[2] 256 parents, all with children aged 4-18, with a spread of those with primary and secondary school children, and a good mix of genders and regional spread across England; 200 teachers, All engaged in teaching full-time (i.e. no TAs or student teachers), equally split across primary and secondary school education and with a good regional spread across England; and 50 head teachers, all employed at head teacher / deputy head teacher / principal level, equally split across primary and secondary school education, and with a good regional spread across England.

- Many of these concerns relate to the emphasis on testing in evaluating teacher performance, which in turn is causing teachers to pass stress on to their students.
- There is general support for a cross-sector debate about testing and assessments.

Regarding the first two hypotheses, responses indicated as follows:

1. Assessments, testing and examinations *are* perceived by parents and teachers alike to be one of the top issues facing the education sector, being mentioned by 31% of respondents, second to teacher workload (36%) and just ahead of funding cuts to schools (30%), but head teachers are slightly less likely to rate tests and examinations as one of their top issues.

2. Spontaneous associations with testing and assessments are much more likely to be negative than positive, with both teachers and parents associating assessments strongly with 'stress' and 'pressure'.

3. Of all tests and assessments, Key Stage 2 and Key Stage 1 SATs tend to be the most negatively perceived. Teachers tend to be more negative about all assessments than parents.

4. When asked what are the specific issues with testing and assessments, teachers in particular articulate concern about the 'culture' of testing and assessments having a negative impact on teachers and students. Parents are less likely than teachers to identify this challenge.

5. Evaluation of teacher performance being too closely tied to tests and assessments is *not* teachers' most 'front of mind' concern when they think about tests and assessments. However, when asked outright, there is very strong agreement among teachers that the evaluation of teacher performance is too closely tied to tests and assessments.

Martin/Thomas to insert here some graphical displays of the data

Discussion points:

1. Are there any surprises in these findings?

2. Do those within your organization/stakeholder group share similar views?

3. What is the most significant message you take from these findings?

Negative attitudes to assessment are not new. They have existed for decades. So, how seriously should one take the expressed negative feelings? One could conclude that these attitudes are inevitable and perhaps inescapable, rather like attitudes to taxation: we don't like paying taxes and they too can be a source of stress and pressure, but we also know that they support a lot of things that are essential and are not going to go away. Adopting such a position, a logical response in the educational context might be to continue to explain and defend robust accountability for

educational outcomes while endeavouring to keep accountability testing and its associated costs to a minimum.

However, it is notable that respondents were of the view that assessment and examinations *per se* are not what most trouble parents and the teaching profession, but rather it is: 1) the emphasis given to accountability testing relative to assessment for learning, and 2) the *kinds* of tests associated with and the uses of test data in current arrangements for evaluating schools and teachers, that most concerns them. This would suggest that negative attitudes are not inevitable but that:

1. a better balance is needed in the emphasis placed on different kinds of assessment and the uses to which assessments are put; and
2. assessment itself must change to meet new demands and circumstances.

Discussion points:

4. Where do you stand in relation to the above two points of view? Is there some truth in both?

5. Do you have a different point of view?

Assessment in support of learning

There is general agreement that when it comes to improving learning and teaching, the most important use of assessment is in that which provides feedback on decisions about what to teach next (the curriculum) and on outcomes of learning and teaching. David Ausubel, the renowned American psychologist once said: "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".[3] Assessment directed to this end is generally referred to as 'formative' assessment.[4]

Meta-analyses (statistical summarization) of literally thousands of research studies indicate that formative assessment or assessment for learning is one of the most powerful of educational interventions.[5] But there are considerable barriers to realizing its transformative potential. Reorienting teachers' professional practice requires consistent support and encouragement and even when these have been present, teachers have generally lacked the tools to sustain effective formative assessment on a daily basis.

Much classroom assessment might be described as being relatively informal and unsystematic. When teachers do assess more systematically, it is invariably for the

---

3 Ausubel, D. P. (1968). Educational psychology, A cognitive view. New York: Holt, Rinehart and Winston, p.vi.

4 Popham, W.J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

5 Black, P. & Wiliam, D (1998). Assessment and Classroom Learning. *Assessment in Education,* 5 (1), 7-71; Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research,* 77 (1), 81–112.

purpose of making judgements and generating evidence to support a final set of assessment grades that will appear on student's end-of-term/year reports. To tap fully into the power of formative assessment, it is necessary for teachers to:

- be clear about which aspect or qualities of learning they wish pupils to develop, in the form of validated 'maps' or learning progressions;
- have a simple and efficient process for collecting, storing and analysing data about their pupils;
- monitor students and their progress on a daily basis using structured observations and assessment tools linked to the objectives of each lesson and integrated into learning activities to minimize interruption to normal classroom routines; and
- use these data as the starting points for teaching explicitly linked to curriculum objectives and tailored to the needs of individual pupils.

The tools required to implement the above have simply not been available until quite recently, and this has made assessment for learning too onerous for the majority of teachers to implement and sustain.

Assuming this to be the case, it follows that assessment providers could do more to assist teachers in undertaking and using formative assessments to improve learning and better personalise teaching to meet the needs of each student.

An example of a tentative start in this area is Pearson's support for progression from 11-16.[6] This arose in response to the removal of National Curriculum levels and the introduction of the Progress 8 Measure. A progression scale, progression maps and classroom assessments have been created to enable teachers to monitor and manage students' progression at KS3 and KS4. A key feature of the progression maps is cross-referencing to EdExcel GCSE specifications, so that a seamless link can be made between formative and summative assessment.

But much more work is required to provide teachers, particularly in Key Stages 1 and 2, with quality tools to enable them to conduct formative assessment while levering the power of new technologies to create efficient and engaging ways of managing data and of assessing a wider range of learning outcomes.

Discussion points:

6. Do you support a greater emphasis on support for assessment for learning?

7. Can you envisage an assessment system in which summative assessment at all levels always builds upon formative assessment?

Assessment for accountability purposes

[6] For further information, see http://www.pearsonschoolsandfecolleges.co.uk/Secondary/Progression-Services/Progression-Services/Progression-Services/ProgressionServices.aspx

Using assessment for accountability purposes is not new. In 1863, the English government implemented a system in which funds received by individual schools depended in part on students' performance in examinations administered by school inspectors. The "payment by results" system was highly controversial, but an important component of the drive to establish a system of public elementary education for all. It remained in place for 30 years.

The current approach to assessment for accountability purposes dates back to the 1988 Education Reform Act, which introduced market-style reforms and stronger accountability, including England's first National Curriculum and national testing. By 1995, national assessment of seven, 11 and 14 year olds in English and mathematics (and science for 11 and 14 year olds) had been introduced. The GCSE examination was reformed and became the main means of assessment for 16 year olds.

Other countries have taken similar directions, particularly the USA, which has also had a long history of assessment for accountability purposes. The *No Child Left Behind* (NCLB) legislation,[7] 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, but one that quickly proved to be problematic.

The challenge for most countries that have used assessment for accountability purposes has been, and continues to be one of striking a balance between validity in terms of assessing a wide range of curriculum outcomes and the need to maximise reliability while minimising testing time. When the balance is not right, there is a real risk of creating perverse incentives that divert attention to the trivial and away from serious objectives and more instructionally relevant uses of assessment.

Referring to the US context, the Gordon Commission gave the following advice to policymakers:

> *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 realize 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 recognize 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.*[8]

---

7 *No Child Left Behind* is a United States Act of Congress that was a reauthorization of the Elementary and Secondary Education Act.

8 The Gordon Commission on the Future of Assessment in Education (2013). *A public policy statement*. Princeton, NJ: The Gordon Commission, pp. 7-8. Retrievable from:

In other words, balance and alignment are critical when it comes to uses of assessment. Accountability needs to be viewed as a multi-level, shared, reciprocal process that is reasonable, effective and promotes a shared trust in the system. Accountabilities should be within the power of the teachers or schools being held to account, and the emphasis should be on improvement from given starting points rather than on status, and on capacity-building requirements in order to deliver improvement.

It is probably true to say that no country as of this point in time has succeeded in implementing an exemplary educational accountability system that is reliant upon a significant component of external testing. But perhaps there may nevertheless be specific lessons to be learnt from practices in other countries that together could form the ingredients of a better system that in turn could foster more positive attitudes towards assessment?

---

http://www.gordoncommission.org/publicationsreports.html

Discussion points:

8. If accountability *per se* is not the problem, then what do you believe to be the chief problems?

9. To what extent do you believe it both possible and desirable to have an accountability system that has the following characteristics?

    - it is the outcome of a negotiated settlement that involves key stakeholder groups and has the support of the profession;
    - It is a system of accountability built around a balanced report card in which one only sees test data in the context of a wide range of other relevant indicators;
    - the assessment programme includes moderated school assessments and covers a wider area of the curriculum and of learning outcomes than just literacy and numeracy;
    - there are strong support and incentives for schools to introduce quality formative assessment and internal systems for tracking student progress;
    - the emphasis is on ongoing improvement rather than a point in time judgement; assessment data are presented in ways that show change over time and take into account measurement error and the impact of student demographics;
    - schools have the opportunity to comment on their data and outline their improvement plans; and
    - test publishers are set the task of working with the profession to come up with state-of-the-art solutions to quality assessment that have greater curriculum validity and that assess the kinds of outcomes that are really important in the modern world.

Looking at international practice

Testing and assessment practices vary widely internationally. To provide an insight into practices in other countries the appendix to this paper provides brief descriptions of accountability testing in five other international jurisdictions relative to those in England, namely: Australia; Canada–Ontario; Finland; Hong Kong, and; USA-Massachusetts. Attention is given to programmes in these countries that assess students in the compulsory years as none include public examinations within their accountability assessment programmes.

All five have performed well in PISA assessments. Australia has slipped over the years but is still in the top 20, Canada performs particularly well, particularly in mathematics, with Ontario (the largest of the Provinces) slightly below the Canadian average. Since the first administration of PISA, Finland was the number 1 performer, but slipped significantly from its leading position in the 2012 results. Hong Kong has

always been in a top position but since 2012 has been somewhat eclipsed by China-Shanghai and with Korea and Japan now scoring at similar levels.

Of the five international jurisdictions, all except Finland have implemented population testing for accountability purposes. In Finland, there is no high-stakes testing programme, although, a schedule (2013-2023) of thematic evaluations is in place, including assessments of learning outcomes, which in turn may involve national sample testing. Teachers have full responsibility for assessment in their respective subjects on the basis of the objectives included in the national curriculum. However, it is also true that Finland has a highly qualified teaching force, with all being required to hold a masters degree.

In the other four, testing programmes similar to the English programme have been implemented with cross-party support and in all jurisdictions assessment is controversial, particularly among teachers, with almost identical criticisms being voiced in each jurisdiction. However, there are some differences when it comes to the mitigation strategies that have been implemented to minimise negative attitudes and backwash effects. These include the following:

- complete avoidance of high stakes testing (Finland being a frequently cited example and an example also of the advantages of having a highly qualified teaching workforce);
- a reduction in the amount of testing or making certain tests optional (with England and Hong Kong being notable examples);
- confining accountability testing to the compulsory years of schooling and excluding senior secondary examinations that serve certification and selection purposes (as in all other jurisdictions except the UK);
- involving parent organisations and teacher unions in the design of the reporting system (as occurred in Australia in the design of the MySchool website);
- limiting consequences to making available individual school results, usually via the internet, along with other relevant contextual information (all jurisdictions other than Finland and the USA);
- no publication of school results, as in Hong Kong, where schools are explicitly required to *not* release their results;
- presenting results with confidence intervals that acknowledge the degree of measurement error associated with estimates of a school's performance (as in Australia and indeed in PISA statistical tables of country performance);
- ensuring a close alignment between the tests and the curriculum (notably in Canada – Ontario, England, Hong Kong and USA – Massachusetts);
- assessing a wide range of curriculum outcomes (as in Australia's Sample Assessment Programme, the use of teacher-led profile assessment in England, Finland's 10-year schedule of thematic evaluations, Hong Kong's testing of Listening, Speaking and Audio-Visual and its use of a number of equated subtests to ensure coverage of a wide range of outcomes within each tested domain);
- giving strong central support to formative classroom assessment (as in Hong Kong's Student Assessment system); and
- moving to online assessment to allow quicker return of results, take advantage of this medium to create more interactive and engaging questions, through adaptive testing to better match tests to students abilities and through matrix sampling to

assess a wider range of curriculum outcomes (notably in Australia, Hong Kong and the USA – Massachusetts).

Discussion points:

10. Accepting there is little evidence of the effectiveness of specific mitigation strategies in creating positive attitudes to assessment in the countries considered, do you think that any of the above do or could contribute to a less negative attitude within the English context?

11. What do you see as the upside and the downside of each of the above strategies?

12. What other strategies might be effective in changing attitudes?

Looking to the future

Earlier, reference was made to ongoing challenges in striking a balance between validity and reliability of assessments. These challenges have been particularly acute when it comes to ensuring the integrity of senior secondary examinations. On the one hand, there is universal acknowledgement that many subjects include subject-specific elements that cannot be effectively assessed through written exams. On the other hand, despite an increasing range of controls over the design and administration of assessments, problems with authenticity of student work and reliability of marking have persisted.[9]

As a consequence, a position has been taken in recent reforms of GCSEs and A-levels that teacher assessment is inherently problematic. Ofqual have therefore prioritised examined assessments in all the most recent developments with non-examined assessment existing only where there is the strongest possible case for its inclusion. This is an understandable and justifiable conclusion, but a disappointing one in that it accepts that assessment of certain important outcomes is currently unrealizable, particularly in high-stakes situations.

At the same time as these certification reforms have been underway, calls for assessment and reporting of a wider range of cognitive, inter- and intra-personal competencies have been gathering momentum. These so-called '21st century skills' include a range of outcomes that are simply not amenable to assessment through written exams, including creativity and innovation, solving complex problems, working collaboratively, metacognition and communication skills.

Herein lies a dilemma. Steps to ensure the integrity of assessment for accountability and certification purposes can run counter to calls to assess outcomes that are as important as those currently assessed. So what of the longer-term future? Are there directions to follow that might ultimately resolve this dilemma?

---

9 See, especially: Ofqual (2013). Review of Controlled Assessments in GCSEs. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/377903/2013-06-11-review-of-controlled-assessment-in-GCSEs.pdf

One path to follow might be to seek to raise professional standards to a level at which trust in the profession is high, perverse incentives are removed and the integrity of assessments is not in doubt. Finland is probably the jurisdiction that has been the closest to exemplifying this situation although consideration of trust within other professions rather than jurisdictions within education is perhaps more informative.

Another path is to look to technological progress for solutions. Among those at the frontiers of thinking about assessment, there is a belief that new technologies, including artificial intelligence and highly interactive online assessment, can open up new possibilities for assessing both examinable and non-examinable outcomes, particularly through building it seamlessly into carefully designed learning tasks completed in an online environment in which there is objective monitoring of the behaviours of learners and a detailed picture of their developing knowledge skills and other attributes.[10]

PISA's initiative in assessing collaborative problem solving is a significant but early step in this direction. For PISA 2015, collaborative problem-solving was defined as "… the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution"[11]. The cognitive and the non-cognitive outcomes embedded in the concept of collaborative problem solving were treated not as made up of separate domains but as a single construct.

The above two directions – changing accountability through increased professionalism and changing assessment through new technologies - are not mutually exclusive and neither is likely to provide a complete answer, but both merit consideration as longer term directions of travel.

Perhaps in the future we can have our cake and eat it: perhaps we can both validly and reliably assess the full range of outcomes that are essential to life, further study and work in the modern world?

Discussion points:

13. Can more be done to resolve tensions between validity and reliability in accountability testing?

14. What, if anything, should we be doing now in terms of promoting research and development into longer-term directions for improving the integrity of assessment for accountability and certification?

15. What are the biggest blockers to progress in this area?

---

10 See, for example, discussion of emerging approaches in a USA context in Conley, D.T. (2004). *A new era for educational assessment*. Retrieved from: http://www.jff.org/sites/default/files/publications/materials/A-New-Era-for-Educational-Assessment-092414_0.pdf

11 OECD (2013). *PISA 2015 Draft Collaborative Problem Solving Framework*. Paris: OECD. Retrieved from: https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

Appendix: Testing and Assessment Programmes in six countries

Australia

Testing programme	National Assessment Program, comprising: • NAPLAN (National Assessment of Literacy and Numeracy) and • NAP Sample Assessments
Testing authority and test developer	ACARA (Australian Curriculum, Assessment and Reporting Authority) on behalf of the Ministerial Council for Education, Early Childhood Development and Youth Affairs (MCEECDYA) representing the Federal and all State and Territory governments.
Commencement	In operation since 2008.
Assessments (Subjects, year levels, frequency)	• Literacy (reading, writing and language conventions) and numeracy administered annually to all students in Years 3,5,7, and 9. • Science literacy, Civics and citizenship, and ICT literacy administered in sample schools on a three-year cycle.
Format of assessments	Mostly multiple choice with some short response questions, but open response for writing
Mode of delivery	• NAPLAN paper and pencil until 2016, but with online delivery to be phased in between 2017-19. • NAP Sample Assessments delivered online.
Use of data and reporting of results	• Detailed statistical information for the nation and for all States and Territories provided in an annual report on Australia's school education sector on agreed key performance measures for schooling including enrolment, attendance, student achievement in national assessments and transitions to further education and work. http://www.acara.edu.au/_resources/20151210_ANR_2012_Parts_1-6_8_and_10.pdf • Publication of eight years of individual school performance for all Australian schools of NAPLAN test results alongside a wide range of other indicators and information on the *MySchool* website. http://www.myschool.edu.au

Australia (contd.)

Mitigation strategies to avoid negative backwash effects	• Consequences limited to publication of a school's NAPLAN results on the *MySchool* website • Engagement of parent organisations and teacher unions in contributing to the design of the website • Active policy to discourage publication of league tables in the media (although not always effective) • Individual school results published with confidence intervals and effect size • Capacity to compare a school's results with those in schools with similar students • Use of Sample Assessment Programme to assess outcomes other than literacy and numeracy
Attitudes to assessment	• Broad acceptance of the assessment programme within the general public, but less positive within the profession. • Criticism tends to focus on the dangers of misleading comparisons and on the limited range of outcomes measured. • Certain pressure groups routinely call for its abolition.
Further information	http://www.acara.edu.au/home

Canada - Ontario

Testing programme	Ontario Provincial Assessment Program • Grades 3 & 6: Assessment of Reading, Writing and Mathematics • Grade 9: Assessment of Mathematics • Grade 10: Ontario Secondary School Literacy Test (OSSLT) - a compulsory standardized test of literacy and a graduation requirement of the Ontario Secondary School Diploma
Testing authority and test developer	Education Quality and Accountability Office (an independent crown agency of the Ontario government)
Commencement	• Grade 3 testing commencing in 1996-97 school year • Grade 6 testing commencing in 1998-99 school year • Grade 9 testing commencing in 2000-01 school year • Grade 10 OSSLT testing commencing in 2000.
Assessments (Subjects, year levels, frequency)	• Reading, writing and mathematics administered to all students in Grades 3 (end of primary division) and 6 (end of junior division) • Mathematics administered to all students in Grade 9 (first year of secondary school) • Literacy (OSSLT) reading and writing administered annually to students in Year 10.
Format of assessments	• Reading: multiple choice and open response • Writing: Long and short writing tasks and multiple choice
Mode of delivery	Paper and pencil
Use of data and reporting of results	• Published report on overall performance to the Minister of education and the general public. • Online reports for individual School Boards and schools via the EQAO website in the form of percentages at or above the Provincial (Ontario) standard, alongside contextual information about the student population and student attitudes to the subjects tested. • Individual Student Reports on each child provided to parents, along with accompanying detailed explanations and suggestions as to how they can support learning, together with summary information on the performance of the school attended by the child and of the School Board. • Data portal for researchers wanting to access EQAO data files.

Canada - Ontario (contd.)

Mitigation strategies to avoid negative backwash effects	• Assessments are curriculum-based and measure achievement of Provincial learning expectations • Educators are involved in all aspects of assessments to ensure they are appropriate, relevant to the curriculum and build the professional capacity of educators in curriculum and assessment • The performance nature of the assessments is maintained by including open-response as well as multiple-choice items (questions)
Attitudes to assessment	• According to the Ontario Institute for Studies in Education at the University of Toronto study Public Attitudes Toward Education in Ontario (2012), 64% of the public supports Province-wide testing at the elementary level and 70% of the public supports Province-wide testing of all students at the secondary level. • Extensive criticism has been expressed from the initiation of the programme by a number of groups, particularly teacher unions, regarding the validity of the tests in terms of reflecting what students have learnt and failure to assess the full range of curriculum outcomes, stress and pressure on students, teachers and parents, the timing of the tests, lack of transparency regarding the measurement properties of the tests and the cost of the testing programme.
Further information	http://www.eqao.com/en

England

Testing programme	National Curriculum assessments
Testing authority and test developer	Developed by the Standards and Testing Agency (STA), an executive agency of the Department for Education, regulated by the examinations regulator, Ofqual
Duration	Introduced between 1991 and1998, but with significant modifications to a national testing from a programme that involved national tests for 7, 11 and 14 year-old students.
Assessments (Subjects, year levels, frequency)	Early Years: • Annual collection of moderated teacher assessments of all students in early years settings using a three-point scale ('emerging', 'expected', 'exceeding') using an EYFS profile covering the following prime areas: communication and language; physical development; and personal, social and emotional development, and the specific areas of literacy, mathematics, understanding the world; and expressive arts and design. Key Stage 1 (Year 2) • Annual national testing of all students in maintained schools in English and mathematics (previously pupils were administered in addition tests of English writing and English grammar, punctuation and spelling). • Teacher assessments of English reading, English writing, mathematics and science using an Interim teacher assessment framework Key Stage 2 (Year 6): • Annual administration to all students in state-maintained schools of national tests in English (Reading; Spelling, grammar and punctuation) and Mathematics, and biennially in Science to a nationally representative sample. • Moderated teacher assessments of English reading, English writing, mathematics and science using an Interim teacher assessment framework
Format of assessments	Teacher-led: Profile assessment using a 3-point rating scale Key Stage 2 tests: Various formats, mostly objective
Mode of delivery	Teacher-led: Frameworks accessed online Key stage tests: Paper-and-pencil

England (contd.)

Use of data and reporting of results	Assessment data for all three key stages are published nationally in performance tables produced by the Department for Children, Schools and Families alongside examination results for secondary schools relating to performance at Key Stage 4
Mitigation strategies to avoid negative backwash effects	• Reduction in the amount of testing • Significant reliance on teacher led profile assessment of a wider range of curriculum objectives
Attitudes to assessment	• Key Stage testing is negatively perceived, particularly by teachers, but also by parents. • Teachers express concern about the 'culture' of testing and assessments having a negative impact on teachers and students. • There is very strong agreement among teachers that the evaluation of teacher performance is too closely tied to tests and assessments.
Further information	https://www.gov.uk/government/organisations/standards-and-testing-agency

Finland

Testing programme	National Evaluation of Primary and Secondary Education There is no high-stakes testing programme. There is, however, a schedule (2013-2023) of thematic evaluations, including assessments of learning outcomes, which in turn may involve national sample testing. During 2016, a frame of reference and an evaluation handbook is being developed as a basis for assessments of learning outcomes in basic education.
Testing authority and test developer	There is no testing authority per se. The only national examination, the matriculation examination, is held at the end of general upper secondary education. For the period of basic education, the focus is on thematic evaluations. Evaluations are the responsibility of the Finnish Education Evaluation Centre (FINEEC) – formed by combining the evaluation activities of the Finnish Higher Education Evaluation Council, the Finnish Education Evaluation Council and the Finnish National Board of Education.
Duration	FINEEC started its operations on 1 May 2014.
Assessments (Subjects, year levels, frequency)	The National Plan for Education Evaluation 2016-2019 indicates that with respect to the evaluation of learning outcomes the focus during this period will be on: • mathematics (grade 9) • Saamai (Lapp), Sign language, Romany • Finnish as a second language
Format of assessments	Not relevant
Mode of delivery	Not relevant
Use of data and reporting of results	Not relevant
Mitigation strategies to avoid negative backwash effects	• No national tests for pupils in basic education in Finland • Complete avoidance of high-stakes testing. • A strong focus on learning rather than testing. • Teachers responsible for assessment in their respective subjects on the basis of the objectives included in the national curriculum.

Attitudes to assessment	• There is a high degree of trust in teacher professional competence. • Teacher and school autonomy are highly regarded. • In recent years, Finland's declining performance in PISA has generated concern and self reflection.
Further information	http://www.oph.fi/english/education_system http://karvi.fi/app/uploads/2016/06/National-Plan-for-Education-Evaluations-2016-2019.pdf

Hong Kong

Testing programme	Basic Competency Assessment, comprising: • TSA (Territory-wide System Assessment) • SA (Student Assessment)[1]
Testing authority and test developer	Hong Kong Examinations and Assessment Authority on behalf of the Education Bureau
Duration	TSA administered since 2004 (Primary 3 since 2004, Primary 6 since 2005 and Secondary 3 since 2006)
Assessments (Subjects, year levels, frequency)	• TSA: Chinese Language (Reading, Writing, Listening Speaking and Audio-Visual), English Language (Reading and Writing, Listening, Speaking) and Mathematics, administered at Primary 3[2], Primary 6 and Secondary 3 (Years 3, 6 and 9) in try-out mode commencing in 2016 with 50 schools participating, being a combination of invitees and volunteers, optional for schools in Primary 6, and all schools annually in Secondary 3. • SA: An online assessment system that assesses students from Primary 1 to Secondary 3 in Chinese Language, English Language and Mathematics covering Chinese, English and Mathematics. It provides instant feedback to students and teachers. Teachers can choose if and when to use the questions or "items" in the web-based bank to construct tests to monitor learning progress.
Format of assessments	• TSA: Students take one of a number of equated sub-papers, thus enabling wider coverage of content but still having scores on a common scale. With the exception of Writing, most items are multiple choice or short response. • SA: Objective questions involving: Click, Highlight, Multiple answer boxes, Drag and drop, and Drop-down lists.
Mode of delivery	• TSA: Paper-and-pencil. • SA: Online, web-based item bank and associated reporting system.
Use of data and reporting of results	• TSA: There is detailed system level report in addition to overall results and trends (including percentages achieving Basic Competence over time), providing an item-by-item analysis of student responses[3]. Reports are also provided to each participating school to help them evaluate the performance of their students and to facilitate the development of plans to improve learning and teaching. None of the reports identify the

	performance of individual students and all school reports are strictly confidential and provided only to the school. • SA: After an assessment is completed, a Student Report is generated showing his/her performance on each item and his/her overall performance. For each item answered incorrectly, the report will show the possible problem leading to that particular wrong answer. There is also a Class Report in the form of a table listing students' answers to highlight overall problems.
Mitigation strategies to avoid negative backwash effects	• Attention paid to assessing the full range of outcomes (including Speaking and Listening) associated with basic competence. • Use of a number of equated subtests to ensure coverage of a wide range of outcomes within each tested domain. • No publication of school results. • Detailed attention to learning from student responses to TSA test questions with follow-up seminars for teachers and detailed item-by-item system and school reports. • Since 2014, no Basic Competency attainment rates for Chinese Language, English Language and Mathematics have been provided to primary schools. • Strong emphasis on formative assessment with schools encouraged to monitor progress through the online Student Assessment system.
Attitudes to assessment	School and teacher views: • Schools have generally supported the retention of TSA but have felt that the mode of reporting could be enhanced with more qualitative data. • In 2015 a small number of schools publicly announced that they would refuse to take the P3 TSA. • The findings from an interview survey conducted by HK Federation of Education Workers with some 600 primary teachers found that: 54% of the interviewees believed that students lost interest in their studies and 45% felt that students had emotional problems due to the TSA. 66% of teachers stated that TSA items were not aligned with students' actual ability levels; 95% of those interviewed stated that TSA was a cause of pressure; 91% of them mentioned that the TSA affected their teaching and planning; 80% found that it was only by drilling that students were able to succeed in the TSA; and 53% thought that the TSA enabled them to better understand students' abilities. Parents' have divided views: • Those who understand the purpose of TSA find that the TSA is not a problem in itself and does not affect the advancement of their children's studies. Yet 'over-

	drilling' by schools seeking to boost TSA results increase pressure on them as parents. • Some suggested that parents should be allowed to have a say on the retention or abolition of TSA. • Some supported the abolition of TSA in the long run. Pressure groups (including Professional Teachers' Union and TSA Concern Group) have negative attitudes. They advocate parents having the right to choose whether their children take part in TSA or not and the abolition of TSA in the long run. The general public has divided views ranging from support for the TSA, to its abolition at all levels. The Education Bureau's Coordinating Committee on Basic Competency Assessment and Assessment Literacy believe that the TSA should continue, but that: • test papers and item designs should be improved to better align them with the requirements of basic competency and tie in with schools' everyday teaching and learning needs, thus, eliminating the need for schools and students to prepare for TSA by drilling; • enhanced training and development for teachers on teaching and assessment literacy should be provided; • further attention should be given to assuring the education sector of the low-stakes nature of the TSA; • efforts to educate the general public about the TSA and to promote assessment literacy should be stepped up; • the emphasis on assessment in support of learning in schools should be strengthened; • transparency should be enhanced through better communication with parents.
Further information	http://www.hkeaa.edu.hk/en/sa_tsa/ http://www.edb.gov.hk/attachment/en/curriculum-development/tsa/fullreport.pdf

USA - Massachusetts

Testing programme	Massachusetts Comprehensive Assessment System (MCAS), including the MCAS Alternate Assessment (MCAS-Alt)
Testing authority and test developer	The Partnership for Assessment of Readiness for College and Careers (PARCC) on behalf of the Massachusetts Department of Elementary and Secondary Education
Duration	The MCAS programme was developed in 1993
Assessments (Subjects, year levels, frequency)	Annual testing as follows: Grade 3 Reading, Mathematics Grades 4, 6 &7 Reading Comprehension & Mathematics Grade 5, 8 &10 Reading Comprehension, Mathematics & Science and Technology/Engineering Note: History and Social Sciences testing in Years 8 and 10 placed on hold due to budgetary concerns
Mode of delivery	In transition from paper-and-paper to onscreen (to be fully implemented by 2019)
Use of data and reporting of results	• The MCAS scale ranges from 200 to 280: Advanced (260-280); Proficient (240-259);Needs Improvement (220-239); Warning/Failing (200-219) • An Educational Proficiency Plan EPP must be developed for the subject matter area(s) in English Language Arts and mathematics in which students did not meet or exceed a scaled score of 240 • MCAS school and district level reports are released each summer on the Department of Elementary and Secondary Education website
Mitigation strategies to avoid negative backwash effects	The development of MCAS 2.0 will focus on: • Combining PARCC, MCAS, and newly developed items into new ELA and math tests • Creating stronger alignment to the Massachusetts Curriculum Frameworks - Introducing writing in response to text at all grades - Developing and implementing new types of items that more deeply assess the standards - Assessing Massachusetts-specific standards • Developing a consistent set of performance standards across all grades • Phasing in computer-based testing

USA – Massachusetts (contd.)

Attitudes to assessment	Criticism tends to be directed at the pressure placed on teachers and students, the tendency for teaching to be restricted to what is tested and the failure of reports to account for student background characteristics.
Further information	http://www.doe.mass.edu/mcas/cal.html https://en.wikipedia.org/wiki/Massachusetts_Comprehensive_Assessment_System