

# Review of Marking Internationally



February 2014

Ofqual/14/5370

# Contents

Introduction.....	2
The assessments included in the review.....	3
Who marks exam scripts? .....	6
Double and multiple marking of assessments .....	9
On-screen marking of assessments .....	14
Quality assurance of the marking process.....	17
Examiner reliability .....	22
Appendix A – Overview of the assessments included in the study.....	24
References .....	29

## Introduction

As part of our standards objective outlined in the Education Act 2011, London: HMSO, we have a requirement to review qualifications available outside the UK. This is to make sure the qualifications we regulate have a level of attainment that is consistent with comparable qualifications taken by students in other jurisdictions.

We have already published our findings from a study of senior secondary assessment in our report *International Comparisons in Senior Secondary Assessment (ICOSSA)*.<sup>1</sup> We have since gathered information and assessment materials from partner jurisdictions for assessments that students take at the end of lower secondary education, at around 16 years of age, to inform future international projects.

This review aimed to give an international perspective to our review of quality of marking by looking at how assessments and tests are marked internationally. This enables us to understand how marking is carried out in other jurisdictions around the world and the importance each jurisdiction places on the quality of their marking.

Our review was based on four research questions:

1. Who marks assessments in other countries?
2. Do any jurisdictions use double and multiple marking? Is there any evidence as to the rationale behind or impact of this? How is it managed, and is it targeted at certain subjects and/or types of questions?
3. Do any jurisdictions use on-screen marking? If so, do they use item-level marking? Is there any evidence as to the impact of this?
4. How do other jurisdictions' quality assurance of exam marking compare with our own? This includes how the quality of the interpretation of the mark scheme is monitored, how clerical errors are eliminated and how inter-rater<sup>2</sup> reliability is monitored and reported.

To allow a direct comparison with the marking of GCSEs and A levels, the primary focus of the review was externally marked qualifications that students take at around age 16 or 18 (the same age as they take GCSEs or A levels at in England). However, to enable us to analyse a wider range of marking systems, the review was extended

---

<sup>1</sup> [www.ofqual.gov.uk/documents/international-comparisons-in-senior-secondary-assessment-summary-report/all-versions](http://www.ofqual.gov.uk/documents/international-comparisons-in-senior-secondary-assessment-summary-report/all-versions)

<sup>2</sup> The degree of agreement among examiners.

to include high-stakes standardised tests and entrance exams taken by students aged 16 to 18 years.

The evidence presented in this review has come from assessment materials and information gathered from our partner jurisdictions, to inform our international studies, and the websites of regional and national education and testing agencies.

### **The assessments included in the review**

The assessments included in this review fall into two main groups: those that award a qualification on completion of a course of study and are standardised tests that act as entrance exams for further study; or reference exams that evaluate education systems and the performance of students at a certain point in their education. Some of the assessments were a combination of both groups.

The assessments and tests we reviewed use a variety of question types including: multiple-choice; structured questions; and open-ended essay questions.

We selected the assessments included based on the following criteria:

- Assessments we were already familiar with from partner jurisdictions that participated in our *International Comparisons in Senior Secondary Assessment* study which could contribute to one or more of the research questions. These assessments are commonly taken by students looking to enter higher education.
  - Australia – New South Wales Higher School Certificate (HSC)
  - Canada – Alberta Diploma
  - Hong Kong – Diploma of Secondary Education (HKDSE)
  - New Zealand – National Certificate of Educational Achievement (NCEA) Level 3
  - People’s Republic of China – Gāokǎo ('National Higher Education Entrance Examination')
  - USA – ACT.<sup>3</sup>

These assessments, with the exception of the USA – ACT and the Gāokǎo, are qualifications that include an element of external assessment with a variety of question types. The ACT is one of the college entrance tests available to high

---

<sup>3</sup> Originally an abbreviation of American College Testing.

school students in the USA. The questions are multiple-choice, but there is also an optional writing test. The Gāokāo is an annual academic exam that is used to select students for university admission. The assessments include a variety of questions types.

- Assessments from additional jurisdictions that have agreed to take part in a proposed study looking at assessments taken by students at the end of the lower secondary phase of their education.
  - Canada – Ontario Secondary School Literacy Test (OSSLT)
  - Chinese Taipei – Basic Competence Test for Junior High School Students (BCTEST)
  - Republic of Korea – National Assessment of Educational Achievement (NAEA) tests
  - USA – Massachusetts Comprehensive Assessment System (MCAS).

The OSSLT and MCAS high school tests are assessments that students must pass to receive the Ontario Secondary School Diploma and the Massachusetts High School Diploma, respectively. The Chinese Taipei BCTEST acts as an entrance exam for students wishing to enter senior high school. The Republic of Korea NAEA is used to identify the rate of progress of students. However, it is also used to evaluate the system and devise school league tables.

All partner jurisdictions taking part in the international studies detailed above were asked to supply specifications, question papers and other materials relating to their assessments. Much of the evidence we present in this review has been taken from these assessment materials and information gathered from our partner jurisdictions.

Because most of the information we already have or have been able to find is for assessments in North America, Australasia and Asia, we carried out an Internet search of European jurisdictions. We found very little relevant information online, so it is likely that the type of information we are looking for is either not published or not published in English by assessment providers in Europe. You can see an overview of the assessments included at appendix A.

Due to the limited amount of information available it was not possible to answer all the research questions for each jurisdiction, and in some cases we are able to describe processes in certain jurisdictions in greater depth than those in others.

The table below gives an overview of the jurisdictions studied when addressing each research question.

**Table 1: The jurisdictions for which data was available to inform each research question**

Jurisdictions included in each section	Who marks exam scripts?	Double marking of assessments	On-screen marking of assessments	Quality assurance of the marking process	Examiner reliability
Australia – New South Wales	✓	✓	✓	✓	
Canada – Alberta	✓				
Canada – Ontario	✓	✓	✓	✓	✓
Chinese Taipei		✓	✓	✓	
Hong Kong	✓	✓	✓	✓	
New Zealand	✓			✓	
People's Republic of China		✓			
Republic of Korea	✓	✓	✓	✓	
USA – Massachusetts	✓	✓	✓	✓	✓
USA – ACT		✓			✓

## Who marks exam scripts?

Our *Review of Quality of Marking in Exams in A levels, GCSEs and Other Academic Qualifications – Interim Report*<sup>4</sup> found that nearly all examiners<sup>5</sup> who mark GCSEs, A levels and other academic qualifications in the UK have considerable teaching experience and subject knowledge. Exam boards in the UK typically only recruit examiners if they have some degree of teaching experience, with most still currently teaching. The rest are former teachers or lecturers. These examiners do not work for exam boards full time and are contracted for a single exam series.

We found that, generally, where the information was available, the jurisdictions we reviewed use a similar model for recruiting examiners to mark assessments. The assessments all included some forms of open response questions, where marking requires expert judgement and a consistent interpretation of the mark scheme.

Most of the jurisdictions we reviewed recruit teachers or retired teachers and require them to have reached a certain level of education and have a good knowledge of the subject they are marking. However, there are some differences. To be eligible to mark in Alberta, teachers must be currently teaching the course they wish to mark (Education Alberta, 2013), whereas, teachers currently under contract or employed in Massachusetts schools are not eligible to mark the MCAS (Massachusetts Department of Elementary & Secondary Education, 2011). In most jurisdictions, prospective examiners apply for positions that are advertised by the exam board or testing agency. An exception is in Alberta, where teachers are nominated by their school superintendent (Education Alberta, 2013).

The HSC in New South Wales consists of a number of different subjects. Depending on the subject, the question types used in exams vary, with multiple-choice, structured response and some open response questions in technical subjects such as maths and science, and open response essay-type questions in English and history. The assessments are marked by experienced teachers, who apply for marking positions and are selected for their experience and subject knowledge. They are required to declare if a family member is studying the HSC (Board of Studies NSW, 2013).

---

<sup>4</sup> [www.ofqual.gov.uk/documents/review-of-quality-of-marking-in-exams-in-a-levels-gcse-and-other-academic-qualifications-interim-report](http://www.ofqual.gov.uk/documents/review-of-quality-of-marking-in-exams-in-a-levels-gcse-and-other-academic-qualifications-interim-report)

<sup>5</sup> Internationally, examiners can be known by other names such as raters or scorers. In this report, we refer to them as examiners.

The Alberta Diploma also consists of a number of different subjects. The question types vary with the subject. Maths and science use multiple-choice and structured response questions, while English and history use open response essay questions. The Diploma assessments are marked by teachers. They are nominated for marking by their superintendent of school or, for private schools, by the principal. To be eligible, a teacher must have taught the course for two or more years (one or more years for English language arts), be teaching the course in the current school year, and have an Alberta Permanent Professional Teaching Certificate (Education Alberta, 2013).

The New Zealand NCEA is made up of credits from a number of different subjects. The assessments include structured and open-ended questions in most subjects. In the assessments we saw, there was no multiple-choice element. The externally marked assessments are marked by teachers or other education professionals contracted to New Zealand Qualifications Authority for that purpose. All teachers that apply for these part-time roles are expected to have curriculum knowledge and teaching experience for the appropriate level, and an understanding of and experience in standards-based assessment (New Zealand Qualifications Authority, 2010).

The Hong Kong HKDSE is made up of a number of subjects. Assessments use a number of question types, with multiple-choice commonly used in subjects such as maths and science as well as structured and open ended questions. Subjects such as English and history use open-ended essay-type questions. The Hong Kong Examinations and Assessment Authority recruits examiners each year to mark the HKDSE. Most are teachers who apply for positions advertised by the Hong Kong Examinations and Assessment Authority. To become an examiner, applicants need to be currently teaching a relevant subject or recently retired from teaching (Hong Kong Examinations and Assessment Authority, 2012a).

The Ontario OSSLT is made up of multiple-choice and essay-type questions. The assessments are marked primarily by teachers. The Education Quality and Accountability Office recruits as many teachers (specifically members of the Ontario College of Teachers) as possible and fills the complement with educators and qualified non-educators. Teachers apply to the Education Quality and Accountability Office, and a representative number are selected based on a set of criteria that include: geographic location covering all areas of the province both urban and rural, current teaching experience, assessment expertise and previous experience with the Education Quality and Accountability Office (Education Quality and Accountability Office, 2011).

The Massachusetts MCAS test has a mix of multiple-choice and open response questions. The test items are scored by trained examiners appointed by the state's assessment contractor. To be eligible to mark the high school MCAS test items, the



examiners must have an undergraduate degree in the subject being marked or in a related area. Teachers and tutors under contract or employed in Massachusetts schools, and anyone under 18 years of age, are not eligible to mark MCAS responses (Massachusetts Department of Elementary & Secondary Education, 2011).

The NAEA tests in the Republic of Korea have a mix of multiple-choice, constructed response and open response questions. The Republic of Korea Institute for Curriculum and Evaluation generally uses examiners who are experienced teachers or postgraduate students in relevant subjects to mark NAEA tests (Korea Institute of Curriculum and Evaluation, 2013).

The jurisdictions reviewed above use a mixture of multiple-choice, structured response and open-ended response items. Like in the UK, the multiple-choice items are often machine marked and the other items are marked by expert examiners.

The expert examiners used in these jurisdictions tend to be teachers or recently retired teachers with expertise in the subject or a related subject. Most apply to become examiners, the exception being in Alberta where examiners must be nominated by their school superintendent.

There are differences in how examiners are selected in different countries. Most are selected based on their teaching and subject experience and suitability, but some jurisdictions place further restrictions on eligibility. For example, teachers, tutors and administrators in Massachusetts who are currently under contract or employed in Massachusetts schools are not eligible to mark. In Ontario, authorities make sure they recruit a representative mix of examiners.

## **Double and multiple marking of assessments**

Double and multiple marking of essay-type, open-ended, constructed response and short answer exam questions is carried out by exam boards and testing agencies for a number of reasons: to ensure the quality of marking; increase the reliability of their marking; and increase confidence in the marking process.

In double marking, two examiners independently assess each student response. The final mark is the combination of two separate marks. In multiple marking, more than two examiners are used. The combination of double or multiple marks to produce a final score is an acknowledgement that legitimate differences in opinion can exist between markers.

Multiple marking is not routinely carried out in GCSEs and A levels. Instead, marking is generally monitored using a sampling approach whereby senior examiners review or re-mark a sample of examiners' allocations.

We found that a number of the international assessments we reviewed routinely use a system of double marking. This is most often used in assessments in which students provide responses that must be judged subjectively, especially where the assessment is high-stakes for the student.

The types of assessments that tend to use double marking include:

- higher education entrance exams (such as the ACT college readiness writing test and the Gāokǎo in China);
- passports to the next stage of education (such as the BCTEST writing test);
- hurdles that must be overcome to gain a qualification (the OSSLT open response reading item and writing task, and all the short answer and open response questions of the high school MCAS test).

The New South Wales HSC routinely uses a double marking system. This is usually applied to questions requiring an extended response, for example essays, creative writing, projects and performances. Short answer questions of the type commonly used in maths and science are not routinely double marked (Board of Studies NSW, 2013). Some double marking is also carried out in Hong Kong, but it is not clear whether this is done routinely or using sampling as part of the quality assurance process (Hong Kong Examinations and Assessment Authority, 2014).

The number of questions or tasks that are double marked in each assessment varies.

The ACT college readiness tests are all multiple-choice apart from the 30-minute writing test (ACT, 2012a). Only the writing test is double marked, the multiple-choice items are machine marked (ACT, 2007).

The BCTEST assesses five subject areas using multiple-choice questions; only the writing test is open response and is double marked (Journal of Educational Research and Development, 2007).

The OSSLT assessment is made up of two papers. The assessments have a mixture of multiple-choice items and open response reading and writing tasks. Only the open response tasks are double marked (Education Quality and Accountability Office, 2012).

The MCAS high school tests for English and maths include multiple-choice and open response items, only the open response items are double marked (Massachusetts Department of Elementary & Secondary Education, 2011).

In the New South Wales HSC, extended response questions are double marked. For the English and history papers provided as part of the ICOSSA study, this means all the questions would be double marked (Board of Studies NSW, 2013).

The number of students taking these assessments and the number of assessments that are marked need to be taken into account when considering double marking. Double marking of question papers can be seen as impractical in assessments where there are very large cohorts. In the summer of 2013 in England, Wales and Northern Ireland, 1,358,477 students took a GCSE unit, 495,610 students took an AS unit and 330,525 students took an A level unit.<sup>6</sup> As students in the UK take multiple GCSEs and A levels, this equates to approximately 5 million GCSEs and 750,000 A levels that need to be marked and over 17 million individual papers.

In contrast, in 2012/13, 134,033 students participated in the OSSLT (Education Quality and Accountability Office, 2013), 210,549 participated in the spring 2012 MCAS high school tests (Massachusetts Department of Elementary & Secondary Education, 2012) and 1.6 million students took the ACT (ACT, 2012b), although as the writing test is optional it is not known how many of these students took the writing test. The New South Wales HSC was taken by 68,409 students in 2011, and 199,198 students took the first BCTEST in 2011 (Research Centre for Psychological and Educational Testing, 2011a). In each case, there are far fewer assessments marked in these jurisdictions compared with England, Wales and Northern Ireland.

---

<sup>6</sup> Figures from the Joint Council for Qualifications. There may be some cross over as some students may have taken GCSE, AS and/or A level units.

The most commonly used double marking model in the systems we reviewed is a blind scoring model. This is where examiners score each response independently: they do not know what mark the second examiner has given the same response.

In Ontario, the Education Quality and Accountability Office's technical report explains that this "ensures that parents, students and teachers can be confident that all students have received valid scores" (Education Quality and Accountability Office, 2012). There are a number of variations of this model, particularly in the way the final mark is arrived at. Some systems use professional judgment to decide which mark should stand. In others, the final mark is derived from the two marks using a set of rules. In the USA, the ACT writing test is marked independently by two examiners and their scores combined. The sum of the scores is the student's writing subscore (ACT, 2007).

In New South Wales, the Board of Studies NSW uses two examiners, who independently judge a student's response. Each examiner allocates a mark using the approved marking guidelines; he or she is unaware of the other examiner's judgement. The final mark is the average of the two. When the marks assigned to a double marked question differ by more than the maximum acceptable difference set by the board, the difference is considered a 'discrepancy' and a third or possibly fourth marking of the student's response takes place. The supervisor of marking or another senior examiner then takes all the independent marks into consideration and uses professional judgement to determine the most appropriate mark for the student's response (Board of Studies NSW, 2013).

In Ontario, the open response reading item and writing task on the high-stakes OSSLT use a blind scoring model. Two examiners independently mark the response, using the same rubric. If the two marks are in exact agreement, that mark is assigned to the student. If the two marks are adjacent, the higher mark (for reading and short writing tasks) or the average of the two marks (for news reports and paragraphs expressing an opinion) is assigned to the student. If the two marks are non-adjacent, the response is marked again by an expert examiner, to determine the correct mark for the student (Education Quality and Accountability Office, 2012).

For the high-stakes Gāokǎo in China, sections of the exam that require individual marking are sent electronically and at random to two examiners for marking. If the variation on the mark is greater than five points, the paper is sent to a third teacher for the final mark (Australian Education International, 2011).

In the USA, the ACT writing test is marked independently by two trained examiners using a six-point scoring rubric. Each examiner rates an essay on a scale ranging from 1 to 6. The sum of the examiners' ratings is a student's writing test subscore on a scale ranging from 2 to 12. During the marking process, differences of more than

one point are evaluated by a third trained examiner to resolve the discrepancies (ACT, 2007).

In Hong Kong, double marked scripts are marked independently by two examiners. If there is a big difference between the two marks, a third examiner will mark the script. If necessary, a second discrepancy marking will take place. The closest pair of marks is added up and this is the mark for the script (Hong Kong Examinations and Assessment Authority, 2012b).

In the Republic of Korea, the NAEA short answer questions are marked using a slightly different model. The questions are marked by pairs of examiners. Each response is independently marked by two examiners and, if they agree, the mark is recorded. If they disagree, the item is passed to a different pair of examiners to mark, and if they agree the mark is recorded. If they disagree, the item is passed to a third pair of examiners. If they agree, the mark is recorded. If they disagree, the item is marked by the Republic of Korea Institute for Curriculum and Evaluation Item Development team and their decision is recorded as the final mark (Korea Institute of Curriculum and Evaluation, 2013).

In Chinese Taipei, a slightly more complex process is used for the writing test that is part of the BCTEST. The test answer sheet for each student is marked by two examiners, randomly chosen. When the difference between the marks given by the two examiners is more than two grades (including two grades), or one of the examiners gives a zero grade, the answer sheet is given to the third examiner to review. The final mark is calculated according to Table 2 below (Research Centre for Psychological and Educational Testing, 2011b).

**Table 2: Scoring of the BCTEST writing test**

	Examiner A	Examiner B	Reviewer	Average score	Final score
<b>Situation 1:</b> Scores the same or one grade difference	5	4		4.5	5
	5	5		5	5
<b>Situation 2:</b> Difference greater than two grades and the reviewer's score is in between the examiners' scores; the final score is determined by the reviewer	5	2	4		4
	5	3	4		4
<b>Situation 3:</b> Difference greater than two grades and the reviewer's score is not in between the scores of	5	2	6	5.5	6

examiners A and B. Final score is the rounded up average of the reviewer's score and the examiner's score closest to it	5	3	1	2	2
<b>Situation 4:</b> If the reviewer's score is not zero, the final score is the average of the reviewer's score and the non-zero examiner score	0	1	0	0	0
	0	2	1	1.5	2
<b>Situation 5:</b> All scores zero; if the reviewer is unable to make a decision, a group decision is made	0	0	0		0
	0	0	?		By decision

For the high-stakes high school MCAS tests, all student responses are marked independently by two examiners. They use a double-blind scoring system, where neither examiner knows whether the response has been marked before and, if it has, what score has been given. A double-blind response with a difference between the two scores (that is a difference greater than one point if there are three or more score points) is sent to the arbitration queue and marked by a senior examiner or quality assurance coordinator. He or she then assigns the final score using the table below (Massachusetts Department of Elementary & Secondary Education, 2011).

**Table 2: Scoring of the MCAS writing test**

Double-blind scoring*			
Examiner 1	Examiner 2	Senior examiner or quality assurance coordinator resolution score	Final score
4	4	-	4
4	3	-	4
3	4	-	4
4	2	3	3
4	1	2	2
3	1	1	1

\* If the examiner scores are identical or adjacent, the highest score is used as the final score.  
If the examiner scores are neither identical nor adjacent, the resolution score is used as the final score.

Our review shows that double marking is frequently used in other jurisdictions. All the assessments we reviewed use double marking in some part, although in most cases only a small number of items are routinely subjected to double marking, as a large proportion of the items are machine marked multiple-choice questions. Writing tasks are the most commonly double marked type of assessment.

However, when comparing the prevalence of double marking internationally with the UK, we must acknowledge the huge difference in the scale of most of these systems. The number of assessments marked in other jurisdictions, when compared with those marked at GCSE and A level, is very small. Therefore, it would require a huge number of additional examiners to apply a double marking model in the same way in the UK.

## **On-screen marking of assessments**

There has been a move in England to the on-screen marking of scripts, including open response and essay-type questions. In general qualifications, around two thirds of all the scripts marked in the summer of 2012 were marked on-screen. On-screen marking is where student scripts are scanned into digital format and marked by human examiners on the computer screen via a secure system.

One of the benefits of on-screen marking is in the administration of the marking process. Scripts no longer need to be physically sent around the country and marks can be automatically totalled and recorded. What's more, the process also allows real-time monitoring of examiners. This allows the quality of marking to be monitored flexibly and continuously, and enables prompt action to be taken when issues arise.

The assessments from the other jurisdictions we reviewed use automated marking systems to mark multiple-choice questions. A number of the assessments that include open response questions are also marked on-screen by human examiners.

In 2012 in Hong Kong, the HKDSE moved from paper-based marking to on-screen marking in public exams. On-screen marking first started in 2007 and now most of the subjects of the HKDSE are marked on-screen. In 2012, the Hong Kong Examinations and Assessment Authority recruited about 3,800 examiners to mark approximately 1.09 million scripts; about 98 per cent of these were marked on-screen (Hong Kong Examinations and Assessment Authority, 2013). New South Wales, the HSC is also moving to on-screen marking. In 2010, approximately 25 per cent of the HSC (and School Certificate) responses were marked on-screen with close to 1,000 teachers taking part in on-screen marking (Board of Studies NSW, 2010). The MCAS high school tests, constructed response items, are marked on-screen (Massachusetts Department of Elementary & Secondary Education, 2011), as are the BCTEST writing test (Journal of Educational Research and Development, 2007) and the NAEA short written answer components (Korea Institute of Curriculum and

Evaluation, 2013). In contrast, the open response questions in the OSSLT are marked on paper (Education Quality and Accountability Office, 2012).

One common theme in the international assessments we reviewed is that marking, whether on-screen or using pen and paper, is generally carried out in marking centres. This is unlike the UK. Marking centres commonly have chief examiners on site to monitor progress, check marked scripts and provide support to examiners. In Hong Kong, examiners can choose from six centres across Hong Kong. Examiners reserve work stations at the assessment centre they wish to work at (Hong Kong Examinations and Assessment Authority, 2012c). In New South Wales, marking is carried out in Sydney venues and in regional marking centres. Different subjects are marked at each centre. In 2012, the Board of Studies NSW indicated that the move to on-screen marking would allow a number of subjects to be marked by examiners at home. However, it is not clear if this now takes place (Lansa, 2014). The MCAS is marked in scoring centres (Massachusetts Department of Elementary & Secondary Education, 2011).

Although the OSSLT is not marked on-screen, the marking is also carried out in marking centres. The examiners enter scores into a personal digital assistant, which uploads the scores to a server to record the marks. This process allows daily data reports to be generated, such as productivity, scoring accuracy and inter-rater reliability (Education Quality and Accountability Office, 2012).

Generally, on-screen marking is approached in two ways: examiners either mark the whole student script, or the scanned script is split into individual questions and they mark individual questions (or items). This is known as item-level marking. Item-level marking allows examiners to become deeply familiar with the mark scheme for that specific item as well as a full range of student answers. It also allows items to be directed to specialist examiners.

Of the assessments that use on-screen marking, most mark at item-level. On-screen marking allows scripts to be readily allocated to different examiners, who can be assigned to mark the answer to a particular question. In Hong Kong, it is not clear how many subjects are marked at item-level, although the Hong Kong Examinations and Assessment Authority gives one of the benefits of on-screen marking as “being flexible allocation of questions to specialist markers” (Hong Kong Examinations and Assessment Authority, 2012a).

In New South Wales, on-screen marking is also carried out at item-level on the HSC. Examiners only see the specific question or section of the exam they are responsible for (Board of Studies NSW, 2010). The NAEA is marked at item-level for all short answer questions for all subjects. The MCAS high school tests are also marked at item-level. The student booklets are scanned into an electronic imaging system. The digitised student responses to constructed response items are sorted into specific



content areas, grade levels and items before being marked. Each individual response is linked to its original answer booklet. This gives the senior examiners access (if they need it) to a student's entire answer booklet (Massachusetts Department of Elementary & Secondary Education, 2011).

Our review found that, like in the UK, other jurisdictions mark on-screen or they are moving towards on-screen marking. The exception was in Ontario, where the OSSLT open response items continue to be marked on paper.

The main difference noted between the UK and the systems reviewed here is that other jurisdictions tend to carry out their marking in marking centres, where the examiners are monitored and they have access to senior examiners. The Board of Studies NSW, as part of its transition process, is moving towards examiners being able to mark at home. However, we should note that if marking were to be carried out in marking centres in the UK, it may be difficult to recruit enough examiners to mark the vast number of papers generated at GCSE and A level.

## **Quality assurance of the marking process**

Making sure marking is carried out to the required standard is a central part of the marking process. It is crucial there are rigorous checks to ensure that the mark schemes are interpreted correctly and consistently by each examiner, and examiners are consistent throughout the marking session. It is also important to make sure there are no clerical errors and marks are added up correctly and assigned to the correct student.

In the UK, all examiners are trained in the application of the mark scheme through a standardisation process. During this process, they practise marking scripts to build up an understanding of the mark scheme and how it should be applied. Standardisation takes place either as face-to-face meetings or, increasingly, online.

Once the standardisation phase has taken place, examiners must qualify to take part in live marking. To do this, they mark a sample of approval scripts or items. These have been pre-marked by a senior examiner and, if an examiner marks his or her scripts within an acceptable tolerance of this mark, he or she is cleared to start live marking.

In the UK during live marking, examiners' work is sampled by senior examiners to check they are applying the mark scheme accurately and consistently. When scripts are marked using the traditional pen and paper method, examiners send senior examiners samples of their marking at a number of agreed points during the process. These scripts are then re-marked to make sure marking is within an acceptable tolerance. The move to on-screen marking allows any sampling checks to be carried out in real time. Senior examiners can identify and resolve any problems at an early stage. In the UK, exam boards use seed scripts or items in each examiner's on-screen batch of scripts to check accuracy. These scripts or items have been given a definitive mark by senior examiners, and examiners must mark within tolerance of this mark.

Some exam boards also back read or spot check samples of examiners' work, effectively re-marking it. Double marking is also carried out on a small sample of scripts by some exam boards as a quality check. If the marks given by the two examiners differ by a given number, a senior examiner will decide what mark to give the work and one or both examiners are given a penalty. The detailed marking process used in GCSEs and A levels does, however, vary by exam board, although the broad process is the same.

In the international assessments we reviewed, we found, where the information was reported, the marking process was similar to England, especially in Hong Kong and New South Wales.

The New South Wales HSC is marked both on-screen and by paper and pen. Examiners attend an examiner briefing, where they are introduced to the marking guidelines, mark schemes and pre-marked benchmark responses. There is opportunity for discussion to ensure that all examiners have a shared understanding of the requirements of the performance levels.

Examiners carry out a pilot session where they apply the marking guidelines. Pilot marking continues until the senior examiner is confident that the examiner is applying the guidelines accurately and consistently. Once marking has started, a sample of the scripts marked by each examiner is check marked by senior examiners to make sure the marks are accurate. If marking guidelines are not being applied correctly then re-briefing takes place. Check marking normally starts with checking a high number of scripts, and this proportion reduces as the examiners' consistency and confidence develops.

In New South Wales, common control scripts are also used to quality check the application of the mark scheme. Generally, a script from each question within the centre is photocopied and distributed to all the examiners of that question at least once per marking session. When on-screen marking is being carried out, the system allows the senior examiner to insert common control scripts at the required frequency. The results on common control scripts are compared with previous sessions and group data to identify inconsistencies in marking. Examiners identified as inconsistent, or as not applying the mark scheme consistently, receive individual attention to redress the problem.

Statistical checks are also used for quality assurance throughout marking and provide a monitoring mechanism for the supervisor of marking. Reports identify: examiners who are marking significantly above or below average for the examiner group; examiners who are using an unusually narrow range of marks; the rate of marking for each member of the examiner group and perceived anomalies between the anticipated target range of marks (as indicated by the Examination Committee's mapping grid); and the spread of marks resulting from the application of the marking guidelines.

In subjects that are single marked, revision marking is carried out as a further quality assurance check. When the marking process is complete, a report is run to identify students whose exam mark is significantly different from their school assessment mark. The responses for these students are re-marked by a group of senior examiners to either confirm or alter the marks (Board of Studies NSW, 2013).

In Hong Kong, examiners attend a meeting where they are briefed on the assessment objectives, the demands of individual questions and how to interpret the mark scheme. They then mark training scripts to familiarise themselves with the mark scheme and the online marking system. They take a qualifying test, which clears

them to take part in live marking. During live marking, the online system randomly distributes scripts with pre-assigned scores to check the standard of the examiners. The system automatically records and calculates marks. Assessment centres have control rooms where the chief examiner and assistant examiners monitor and carry out real-time checks on examiners and take immediate action if a problem is identified.

In Hong Kong, the scripts of each examiner, when marking on paper, undergo at least two stages of check marking to ensure consistency. If a script has been check marked, the mark awarded by the check examiner is also recorded and may override the original mark, if deemed more reliable. A team of checkers also check the addition of marks on all scripts and that all the pages have been marked (Hong Kong Examinations and Assessment Authority, 2012c).

The North American jurisdictions Ontario and Massachusetts have similar processes that are both detailed and tightly controlled. The Ontario OSSLT is marked in scoring rooms. Examiners for open response items undergo training to develop a common understanding of the scoring material. Following this, examiners take a test where they must have at least a 70 per cent exact match with the expertly assigned scores. An examiner who fails the qualifying test twice is dismissed.

Examiners receive on-going training in both morning and afternoon sessions to refresh their understanding of the scoring materials and to ensure that they apply this consistently from one day to the next, and before and after lunch breaks.

Marking leaders begin each day with a review of all or a portion of the rubrics and anchors (sample responses), to refocus examiners and highlight any section of the rubrics that require attention. The review is more comprehensive after a weekend break. Each afternoon, examiners begin by marking one or more of the selected calibration items. The calibration item scores are reviewed, and, if any issues are raised, examiners are provided with an explanation and clear information and guidance on the correct way to score the items. Marking leaders and supervisors receive daily data reports showing daily and cumulative validity, reliability and productivity data for individual examiners and groups of examiners in their room (Education Quality and Accountability Office, 2012).

In Massachusetts, examiners are trained in marking centres. They are trained on a single item and then complete a qualifying test to demonstrate that they can mark that item accurately. Once they have passed the qualifying test, they independently mark student responses to that item. Once the marking of that particular item is complete, they are then trained and qualify to mark subsequent items.

Examiners have to complete recalibration assessments at the beginning of every marking session to qualify to mark. Once they have met or exceeded the minimum

standard on the qualifying set of scripts, they are allowed to begin marking. Examiners are constantly monitored throughout the entire marking window to make sure they score student responses as accurately and consistently as possible. If an examiner falls below the minimum standard on any of the quality control tools, there is some form of intervention, ranging from feedback to retraining to dismissal.

Further quality checks are carried out by the senior examiner. Embedded committee-review responses (responses approved by the chief examiner) are loaded into the system and randomly sent to examiners. Double-blind marking also takes place and random items are selected to carry out read-behind marking. If examiners do not meet the accuracy standard, they are monitored more closely. If an examiner's statistics have not reached the minimum acceptable level by the end of the marking session, his or her scores for that day are voided and the responses returned to the scoring queue (Massachusetts Department of Elementary & Secondary Education, 2011).

We found less detailed information available for other jurisdictions, although the processes they follow appear to be similar to those previously detailed.

In Chinese Taipei, examiners of the BCTEST writing test receive training on the application of scoring rubrics. During the training, they mark 20 trial test papers. They then have to pass a test consisting of three trial papers each time they enter the marking system. They have to repeat this process each time they leave the system for more than 30 minutes. The system automatically inserts random trial test papers to check quality (Journal of Educational Research and Development, 2007).

In New Zealand, all examiners must be trained by their panel leader. Examiners mark the first batch of booklets and record their marks on the electronic system. They send a sample of marking for check marking to the panel leader, who advises on adjustments to individual marking. Once they receive approval to proceed to live marking, the examiners continue marking and send regular samples for checking. The marks entered on the electronic system allow the panel leader to monitor the distribution profile of examiners and advise on adjustments (New Zealand Qualifications Authority, 2013).

In the Republic of Korea, all examiners that mark the NAEA tests are trained, and each examiner must mark 30 scripts with a reliability of 0.8 to be approved for live marking. All the marking sessions are overseen by the Republic of Korea Institute for Curriculum and Evaluation. This allows it to intervene and take action should any issues be identified (Korea Institute of Curriculum and Evaluation, 2013).

All the quality assurance processes reviewed here are broadly similar. They all require examiners to attend training and pass a test to qualify to mark.

Where on-screen marking is used, most jurisdictions continuously generate statistics to monitor examiners. These are regularly reviewed by senior examiners, allowing them to intervene immediately if issues occur. For example, in New South Wales, senior examiners have a real-time dashboard view. This enables them to look at the marking of any script, view statistics such as the distribution of marks and the speed of examiners, and exchange messages with examiners. The chief examiner in Hong Kong monitors examiners in real time from his or her control room. The Massachusetts system continuously monitors examiners using responses approved by the chief examiner to check examiner accuracy. It also carries out double-blind marking, where responses are marked independently by two examiners, and read-behind marking, where responses that have already been marked are re-marked blind by the senior examiner.

## Examiner reliability

In 2008, we launched our Reliability Programme to investigate the reliability of results in national tests, public exams and other qualifications in England. It was set up to help us develop our regulatory policy on reliability and indicate where to improve the assessment systems in use. We commissioned a series of research projects into various aspects of reliability. The reports were published together in *Ofqual's Reliability Compendium*.<sup>7</sup>

One aspect of the programme was investigating inter-rater reliability. A research project by Bramley and Dhawan (2010) looked at examiner-related variability as an aspect of qualification reliability. The project reported details of both paper-based and on-screen marking studies, using GCSE and A level data from one exam board. The introduction to the project report concluded that “marker agreement is fairly high: the data seems to suggest that paper-based marking of public examinations has become more reliable over the years, and marker agreement for on-screen marking, albeit on papers with fewer subjective questions to mark, is impressively high”.

In the UK, exam boards do not routinely publish statistics on the reliability of their marking. This appears to be in line with other jurisdictions. However, they do carry out analysis as part of their quality assurance processes.

All the jurisdictions we reviewed generate various statistics during marking. Generally, these statistics are used during marking to monitor examiners. Only a small number of jurisdictions publish these statistics in exam reports, as evidence of the reliability of their assessments. Where reported, these statistics are often presented in terms of the percentage of examiner agreement on double marked items. The Education Quality and Accountability Office publishes inter-rater reliability estimates for both the reading and writing tasks of the OSSLT in its technical report each year. It cites consistency and accuracy of marking as an important factor in achieving classification accuracy of 0.87 to 0.93 for the OSSLT in 2010/11 (Education Quality and Accountability Office, 2012).

The Education Quality and Accountability Office defines inter-rater reliability as the percentage of agreement between the scores awarded by the pair of examiners on double marked questions. Four indices are used to identify the inter-rater reliability: percentage of exact agreement; percentage of exact-plus-adjacent agreement; percentage of adjacent agreement; and percentage of non-adjacent agreement. The percentage for all four indices is reported for each item in the test, for both the

---

<sup>7</sup> [www.ofqual.gov.uk/standards/research/reliability/compendium](http://www.ofqual.gov.uk/standards/research/reliability/compendium)

reading and writing components of the OSSLT (Education Quality and Accountability Office, 2012).

The Massachusetts Department of Elementary & Secondary Education also publishes inter-rater reliability statistics in its annual MCAS technical reports.

The statistics published are the results of the double-blind marking that is carried out during the marking process. The statistics are presented as evidence of the reliability of the MCAS tests. The statistics show: the number of score categories; the number of included scores; the percentage of exact agreement; percentage of adjacent agreement; correlation between the first two sets of scores; and the percentage of responses that needed a third score (Massachusetts Department of Elementary & Secondary Education, 2011).

Some jurisdictions carry out reliability studies for research purposes to check the reliability of their exams. In the USA, the 2009 ACT writing test technical report outlined a study that was conducted on the ACT writing test. This was a special administration of the test in September 2003. Students took two forms of the test under standardised and secure conditions on consecutive days. Using the data collected, the rater-agreement reliability for the test was estimated using multiple pairs of raters and ranged from 0.92 to 0.94 (ACT, 2009). Examiner reliability is seen as an important aspect of the quality assurance process for exams in all the jurisdictions we reviewed. However, as in the UK, examiner reliability statistics are rarely routinely published internationally.



## Appendix A – Overview of the assessments included in the study

### Australia – New South Wales Higher School Certificate (HSC)

Total students in 2009: 68,409; achieved: 63,518 (Board of Studies NSW, 2011).

The HSC is awarded to students who successfully complete senior high school level studies (Years 11 and 12, or equivalent). It is developed, managed and awarded by the Board of Studies NSW. Typically, the qualification is studied over two years by students aged 17 to 19.

To be eligible to study for the HSC, students must first complete the New South Wales School Certificate. This is awarded by the Board of Studies NSW to eligible students at the end of Year 10.

All courses in the HSC have a unit value. Most courses are worth two units. To gain an HSC, students “must have completed a minimum of 12 units of Preliminary courses and ten units of HSC courses. Students must satisfactorily complete the Preliminary course (usually studied during Year 11) before they are eligible to commence the corresponding HSC course (usually studied during Year 12)” (Board of Studies NSW, 2012).

Students can study a great number of possible courses in a wide range of subject areas. English is the only compulsory component of the HSC.

### Canada – Alberta Diploma

Total students in 2009: 45,478; achieved: 31,332 (Government of Alberta, 2012).

The Alberta Diploma is the main qualification awarded to students completing their high school education in Alberta. It is developed and administered by the Government of Alberta. To gain the Alberta Diploma, students take part in the Grade 12 Diploma Examinations programme. This takes place at the end of the three-year course, when students are typically aged 17 to 18. Established in 1984, the Diploma Examinations programme has three main purposes:

- “to certify the level of individual student achievement in selected Grade 12 courses;
- “to ensure that province-wide standards of achievement are maintained;
- “to report individual and group results” (Government of Alberta, 2011).

### **Canada – Ontario Secondary School Literacy Test (OSSLT)**

In 2012-13, the total cohort eligible was: 143,358; 134,033 fully participated; 110,162 were successful (Education Quality and Accountability Office, 2013).

All students enrolled in Grade 10 in all publicly funded schools in Ontario, as well as in inspected private schools, take the OSSLT, as it is a graduation requirement for all students who wish to receive the Ontario Secondary School Diploma (OSSD).

The OSSLT is administered annually and assesses Grade 10 students' literacy skills based on reading and writing curriculum expectations across all subjects in the Ontario Curriculum up to the end of Grade 9 (Education Quality and Accountability Office, 2012).

### **Chinese Taipei – Basic Competence Test for Junior High School Students (BCTEST)**

In 2011, 199,198 students took the BCTEST (Research Centre for Psychological and Educational Testing, 2011a).

Developed and administered by The Centre for Research and Development of Psychological and Educational Testing at National Taiwan Normal University, BCTEST is a standardised test to measure the educational achievement of junior high school graduates in five subject areas: Chinese, English, maths, natural sciences and social studies. It is used as the entrance test for students wishing to enter senior high school. Students can choose to sit the BCTEST once or twice a year. They can then choose to use either their results from one test or their best results for entrance applications (Journal of Educational Research and Development, 2007).

### **Hong Kong – Diploma of Secondary Education (HKDSE)**

In 2013, 82,283 students registered (Hong Kong Examinations and Assessment Authority, 2013b).

The first HKDSE exams took place in 2012, after the students had completed six years of secondary education. The HKDSE replaced the Hong Kong A level (HKALE) and is aimed at all students. Like the HKALE, it will serve as the university admission exam under the Joint University Programmes Admissions System (JUPAS).

As part of the HKDSE, school students are typically required to take four core subjects (Chinese language, English language, liberal studies and maths), as well as two or three elective subjects (Hong Kong Examinations and Assessment Authority, 2012d).

### **New Zealand – National Certificate of Educational Achievement (NCEA) Level 3**

Total students in 2009: 36,371; achieved: 22,000 (New Zealand Qualifications Authority, 2010).

The NCEA functions as New Zealand's main secondary school qualification. Introduced between 2002 and 2004, it grew out of a long-term intention to "establish standards for national qualifications and recognise a wider range of skills and knowledge" (New Zealand Qualifications Authority, 2012). The Ministry of Education is responsible for the development of the NCEA and the New Zealand Qualifications Authority is responsible for its delivery.

There are three levels of NCEA: 1, 2, and 3. At each level, students must achieve a minimum of 80 credits to gain an NCEA. Generally, students complete Level 1 standard subjects in Year 11, Level 2 in Year 12, and Level 3 in Year 13, although credits can be gained over more than one year. In most cases, 1 credit represents ten hours of learning and assessment. A typical course generates between 18 and 24 credits – so, over five subjects, a student could aim for up to 120 credits.

### **People's Republic of China – Gāokǎo (National Higher Education Entrance Examination)**

Total students in 2010: 9.57 million (Xinhuanet.com 2010).

The Gāokǎo is an annual academic exam used for selecting students for university admission. This exam is a prerequisite for entrance into almost all higher education institutions at undergraduate level. It is usually taken by students in their last year of high school, although there are no age restrictions on entry. To take part in the entrance exam, students must have a senior middle school graduation certificate (Gaozhong). Students with a vocational middle school certificate are officially also allowed to take the Gāokǎo.

Gāokǎo exams are set by provincial-level exam authorities under the control of the National Education Examinations Authority.

As part of the Gāokǎo, students are required to take exams in a range of subjects over a three-day period. In most provinces, students take the three core subjects – Chinese, a foreign language (generally English) and maths – and either the humanities suite (geography, history and politics), or the science suite (biology, chemistry and physics).

### **Republic of Korea – National Assessment of Educational Achievement (NAEA) tests**

There are approximately 600,000 students in each NAEA cohort (Korea Institute of Curriculum and Evaluation, 2013e).

The NAEA tests are administered by The Korean Institute for Curriculum and Evaluation. They are taken by all students in Grades 6, 9 and 10. The tests assess students' achievement levels on the contents of the National Curriculum.

The main purpose of assessment is to identify the rate of progress of students. However, it is also used to evaluate the education system and devise school league tables. Some school districts use NAEA scores to inform their allocation of upper secondary placements.

The subjects tested are: English, Korean, maths, science and social studies.

### **USA – Massachusetts Comprehensive Assessment System (MCAS)**

Some 210,549 students participated in the spring 2012 MCAS high school tests (Massachusetts Department of Elementary & Secondary Education, 2012).

The MCAS is designed to meet the requirements of the Education Reform Law of 1993. This law specifies that the testing program must:

- test all public school students in Massachusetts, including students with disabilities and English language learner students;
- measure performance based on the Massachusetts Curriculum Framework learning standards;
- report on the performance of individual students, schools, and districts.

The Education Reform Law requires students to pass the Grade 10 tests in English language arts, maths and one of the four high school science and technology/engineering tests as one condition of eligibility for a high school diploma (in addition to fulfilling local requirements).

In addition, the MCAS program is used to hold schools and districts accountable, on a yearly basis, for the progress they have made towards the objective of the No Child Left Behind Law that all students be proficient in reading and maths by 2014.

The MCAS tests: English language arts, maths, science and technology/engineering (Massachusetts Department of Elementary & Secondary Education, 2008).

### **USA – ACT**

Total students in 2011: 1.6 million (ACT, 2012a).

The ACT (originally an abbreviation of American College Testing) is a “curriculum- and standards-based educational and career planning tool that assesses students' academic readiness for college” (ACT, 2012a). It is one of the college entrance tests

available to high school students in the USA. The test assesses high school students' general educational development and their readiness for entry-level college course work. The ACT is made up of a multiple-choice test and an optional writing test. The multiple-choice components cover college readiness skills in four content areas: English, maths, reading, and science. The ACT writing test measures skill in planning and writing a short essay (ACT, 2012a).

The content covered by each of the five tests "is drawn from the domain of each content area that educators agree is important to that content area and that is prerequisite to successful performance in entry-level college courses" (ACT, 2010).

## References

- ACT (2007) *The ACT Technical Manual*. Iowa City, ACT.  
Available at: [www.act.org/qualitycore/pdf/TechnicalManual.pdf](http://www.act.org/qualitycore/pdf/TechnicalManual.pdf) (accessed 29th January 2014).
- ACT (2009) *ACT Writing Test Technical Report*. Iowa City, ACT.  
Available at: [www.act.org/research-policy/policy-publications](http://www.act.org/research-policy/policy-publications) (accessed 29th January 2014).
- ACT (2010) *Your Guide to the ACT*. Iowa City, ACT.
- ACT (2012a) *Fact Sheets – The ACT Test*. Iowa City, ACT.  
Available at: [www.act.org/newsroom/factsheets/act.html](http://www.act.org/newsroom/factsheets/act.html) (accessed 23rd January 2013).
- ACT (2012b) *Assessing Academic Growth for College and Career Readiness*. Iowa City, ACT. Available at: [www.act.org/aap](http://www.act.org/aap) (accessed 7th March 2012).
- Australian Education International (2011) *Research on China's National College Entrance Examination (the Gaokao) (2009)*. Canberra, Australian Education International. Available at: [http://sydney.edu.au/ab/committees/admissions/2011/AEI\\_Gaokao\\_Report.pdf](http://sydney.edu.au/ab/committees/admissions/2011/AEI_Gaokao_Report.pdf) (accessed 29th January 2014).
- Board of Studies NSW (2010) *Marking HSC exams onscreen*. [no place], [no publisher]. Available at: <http://news.boardofstudies.nsw.edu.au/index.cfm/2010/11/18/Marking-HSC-exams-onscreen> (accessed 30th August 2013).
- Board of Studies NSW (2011) *Statistics Archive: 2009 Higher School Certificate Total Candidature*. [no place], [no publisher]. Available at: [www.boardofstudies.nsw.edu.au/ebos/static/ebos\\_stats.php](http://www.boardofstudies.nsw.edu.au/ebos/static/ebos_stats.php) (accessed 13th January 2012).
- Board of Studies NSW (2012) *Information for employers about the HSC*. [no place], [no publisher]. Available at: [www.boardofstudies.nsw.edu.au/hsc-results/employers.html](http://www.boardofstudies.nsw.edu.au/hsc-results/employers.html) (accessed 2nd March 2012).
- Board of Studies NSW (2013) *HSC marking*. [no place], [no publisher]. Available at: [www.boardofstudies.nsw.edu.au/hsc\\_exams/marking.html](http://www.boardofstudies.nsw.edu.au/hsc_exams/marking.html) (accessed 3rd September 2013).

Education Alberta (2013) *Information for Markers/Teachers*. [no place], [no publisher]. Available at: <http://education.alberta.ca/admin/testing/diplomaexams/teacherinvolvement/markers.aspx> (accessed 15th August 2013).

Education Quality and Accountability Office (2011) *Technical Report for the 2009–2010 Assessments*. Toronto, Education Quality and Accountability Office. Available at: [www.eqao.com/pdf\\_e/11/Ctr\\_2009-10Report\\_ne\\_0511\\_web.pdf](http://www.eqao.com/pdf_e/11/Ctr_2009-10Report_ne_0511_web.pdf) (accessed 29th January 2014).

Education Quality and Accountability Office (2012) *Technical Report for the 2010–2011 Assessments*. Toronto, Education Quality and Accountability Office. Available at: [www.eqao.com/Publications/ArticleReader.aspx?Lang=E&article=b11A004&section](http://www.eqao.com/Publications/ArticleReader.aspx?Lang=E&article=b11A004&section) (accessed 29th January 2014).

Education Quality and Accountability Office (2013) *Highlights of the Provincial Results: Grade 9 Assessment of Mathematics and Ontario Secondary School Literacy Test (OSSLT) English-Language Students, 2012–2013*. Toronto, Education Quality and Accountability Office. Available at: [www.eqao.com/pdf\\_e/13/Secondary\\_Highlights\\_2013\\_en.pdf](http://www.eqao.com/pdf_e/13/Secondary_Highlights_2013_en.pdf) (accessed 29th January 2014).

Government of Alberta (2011) *General Information Bulletin – Introduction to the Diploma Examinations Program*. [no place], [no publisher]. Available at: [http://education.alberta.ca/media/6446502/02-dip-qib-2011-12\\_intro%20revisions.pdf](http://education.alberta.ca/media/6446502/02-dip-qib-2011-12_intro%20revisions.pdf) (accessed 13th January 2012).

Government of Alberta (2012) *High School Completion Rates - Measure History* [no place], [no publisher]. Available at: [http://ideas.education.alberta.ca/media/78938/apori\\_hsc\\_rates.pdf](http://ideas.education.alberta.ca/media/78938/apori_hsc_rates.pdf) (accessed 5<sup>th</sup> February 2014).

Hong Kong Examinations and Assessment Authority (2012a) *Marking and Grading of Public Examinations*. Hong Kong, Hong Kong Examinations and Assessment Authority. Available at: [www.hkeaa.edu.hk/en/resources/leaflets/](http://www.hkeaa.edu.hk/en/resources/leaflets/) (accessed 29th January 2014).

Hong Kong Examinations and Assessment Authority (2012b) *Hong Kong Diploma of Secondary Education Examination Marking and Processing of Marks July 2012*. Hong Kong, Hong Kong Examinations and Assessment Authority. Available at: [www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/leaflet\\_MPM\\_final.pdf](http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/leaflet_MPM_final.pdf) (accessed 29th January 2014).

Hong Kong Examinations and Assessment Authority (2012c) *Onscreen Marking System*. [no place], [no publisher]. Available at: [www.hkeaa.edu.hk/DocLibrary/HKDSE/text\\_transcript/Onscreen\\_Marking\\_System.pdf](http://www.hkeaa.edu.hk/DocLibrary/HKDSE/text_transcript/Onscreen_Marking_System.pdf) (accessed 29th January 2014).

Hong Kong Examinations and Assessment Authority (2012d) *Video – Onscreen Marking System (transcript)*. Available at: [www.hkeaa.edu.hk/DocLibrary/HKDSE/text\\_transcript/Onscreen\\_Marking\\_System.pdf](http://www.hkeaa.edu.hk/DocLibrary/HKDSE/text_transcript/Onscreen_Marking_System.pdf) (accessed 29th January 2014).

Hong Kong Examinations and Assessment Authority (2013) press release: *2013 Hong Kong Diploma of Secondary Education (HKDSE) Examination Results Released*. Available at: [www.hkeaa.edu.hk/DocLibrary/Media/PR/20130714\\_HKDSE\\_Results\\_ENG\\_FULL.pdf](http://www.hkeaa.edu.hk/DocLibrary/Media/PR/20130714_HKDSE_Results_ENG_FULL.pdf) (accessed 29th January 2014).

Hong Kong Examinations and Assessment Authority (2014) *Marking Procedures for Written Papers. (Category A subjects)*, [no place], [no publisher]. Available at: [http://www.hkeaa.edu.hk/en/HKDSE/assessment/marketing/marketing\\_procedure\\_for\\_written\\_papers/](http://www.hkeaa.edu.hk/en/HKDSE/assessment/marketing/marketing_procedure_for_written_papers/) (accessed 29th January 2014).

Journal of Educational Research and Development (2007) *Review and Prospect of the Basic Competence Test for Junior High School Students* (issue 4, volume 3). Academe Research Journals.

Korea Institute of Curriculum and Evaluation (2013). Meeting of the Korea Institute of Curriculum and Evaluation with Ofqual on 22nd May 2013. Coventry.

Lansa (2014) *OBOS scores top marks for Onscreen Marking*. Available at: <http://news.boardofstudies.nsw.edu.au/index.cfm/2010/3/22/onscreen-marking> (accessed 29th January 2014).

Massachusetts Department of Elementary & Secondary Education (2008) *Ensuring Technical Quality: Policies and Procedures Guiding the Development of the MCAS Tests*. Maldon, Massachusetts Department of Elementary & Secondary Education. Available at: [www.doe.mass.edu/mcas/tech/technical\\_quality.pdf](http://www.doe.mass.edu/mcas/tech/technical_quality.pdf) (accessed 29th January 2014).

Massachusetts Department of Elementary & Secondary Education (2011) *2011 MCAS and MCAS-Alt Technical Report*. Maldon, Massachusetts Department of Elementary & Secondary Education. Available at: [www.mcasservicecenter.com/documents/MA/Technical%20Report/2011/2011%20MCAS%20and%20MCAS%20Alt%20Tech%20Report\\_Final.pdf](http://www.mcasservicecenter.com/documents/MA/Technical%20Report/2011/2011%20MCAS%20and%20MCAS%20Alt%20Tech%20Report_Final.pdf) (accessed 29th January 2014).



Massachusetts Department of Elementary & Secondary Education (2012) *MCAS Tests: Summary of State Results*. Maldon, Massachusetts Department of Elementary & Secondary Education. Available at:

[www.doe.mass.edu/mcas/results.html?yr=2012](http://www.doe.mass.edu/mcas/results.html?yr=2012) (accessed 29th January 2014).

New Zealand Qualifications Authority (2010) *Annual Report on NCEA & New Zealand Scholarship Data & Statistics (2009)*. [no place], [no publisher]. Available at:

[www.nzqa.govt.nz/assets/About-us/Publications/stats-reports/ncea-annualreport-2009.pdf](http://www.nzqa.govt.nz/assets/About-us/Publications/stats-reports/ncea-annualreport-2009.pdf) (accessed 29th January 2014).

New Zealand Qualifications Authority (2012) *Factsheet # 1: The Development of NCEA*. [no place], [no publisher]. Available at: [www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/the-facts/factsheet-1](http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/the-facts/factsheet-1) (accessed 6th March 2012).

New Zealand Qualifications Authority (2013) *NZQA Examination Marking Process*. [no place], [no publisher]. Available at: [www.nzqa.govt.nz/assets/Qualifications-and-standards/Qualifications/NCEA/NZQA-Examination-Marking-Process.pdf](http://www.nzqa.govt.nz/assets/Qualifications-and-standards/Qualifications/NCEA/NZQA-Examination-Marking-Process.pdf) (accessed 29th January 2014).

Research Centre for Psychological and Educational Testing (2011a) BC test 2011 first PR analysis. [no place], [no publisher]. Available at:

[www.bctest.ntnu.edu.tw/10001pr.html](http://www.bctest.ntnu.edu.tw/10001pr.html) (accessed 29th January 2014).

Research Centre for Psychological and Educational Testing (2011b) 評分者，兩個恰恰好？——以國民中學學生基本學力測驗寫作測驗為例 (*Scores, Are Two All Right?*).

Available at: [www.rcpet.ntnu.edu.tw](http://www.rcpet.ntnu.edu.tw) (accessed 29th January 2014).

Xinhuanet.com (2010) news story: *China gears up for annual national college entrance exam*. Available at: [http://news.xinhuanet.com/english2010/culture/2010-06/03/c\\_13330217.htm](http://news.xinhuanet.com/english2010/culture/2010-06/03/c_13330217.htm) (accessed 29th January 2014).

We wish to make our publications widely accessible. Please contact us if you have any specific accessibility requirements.

First published by the Office of Qualifications and Examinations Regulation in 2014

© Crown copyright 2014

You may re-use this publication (not including logos) free of charge in any format or medium, under the terms of the [Open Government Licence](#). To view this licence, visit [The National Archives](#); or write to the Information Policy Team, The National Archives, Kew, Richmond, Surrey, TW9 4DU; or email: [psi@nationalarchives.gsi.gov.uk](mailto:psi@nationalarchives.gsi.gov.uk)

This publication is also available on our website at [www.ofqual.gov.uk](http://www.ofqual.gov.uk)

Any enquiries regarding this publication should be sent to us at:

Office of Qualifications and Examinations Regulation	
Spring Place	2nd Floor
Coventry Business Park	Glendinning House
Herald Avenue	6 Murray Street
Coventry CV5 6UB	Belfast BT1 6DN

Telephone 0300 303 3344

Textphone 0300 303 3345

Helpline 0300 303 3346