



**Evidence for
Excellence in
Education**



This report has been commissioned by the Office of Qualifications and Examinations Regulation.

Report

Maintaining qualification and assessment standards: summary of international practice

National Foundation for
Educational Research
(NFER)



Maintaining qualification and assessment standards: summary of international practice

Newman Burdett
Emily Houghton
Claire Sargent
Jo Tisi

First published by the Office of Qualifications and Examinations Regulation in 2013.

© Crown Copyright 2013

Office of Qualifications and Examinations Regulation
Spring Place
Herald Avenue
Coventry Business Park
Coventry
CV5 6UB

How to cite this publication:

Burdett, N., Houghton, E., Sargent C. and Tisi J. (2013). *Maintaining Qualification and Assessment Standards: Summary of International Practice*. Slough: NFER.



Contents

1	Executive Summary	1
	1.1 Background and purpose of the report	1
	1.2 Research aims	1
	1.3 Findings	2
2	Introduction	4
3	Methodology	6
	3.1 Research overview	6
	3.2 Policy scoping	6
	3.3 Literature review	7
	3.4 Synthesis	7
4	Findings	8
	4.1 The importance of context	8
	4.2 Establishment of standards	8
	4.3 Maintenance of standards	11
	4.4 Reference testing	22
	4.5 Post-examination reviews	22
5	Conclusion	24
	References	25
	Appendix A – Summaries of jurisdiction education systems	33
	Alberta, Canada	36
	Finland	43
	France	48
	Germany	52
	Hong Kong	58
	Massachusetts - USA	68
	The Netherlands	74
	New Zealand	80
	Singapore	88
	South Korea	93
	Appendix B – Reference summaries	100
	Appendix C – Search strategy	104





1 Executive Summary

1.1 Background and purpose of the report

Setting and maintaining standards is a technically challenging issue and one that attracts a lot of public interest. The way standards are maintained in England has come under intense scrutiny and has stood up remarkably well to that scrutiny. This, however, does not mean that much cannot be learnt from studying other systems – even if the final answer to posing the question “What can we learn from how other countries establish and maintain standards?” is that England uses methods that are the same as or as good as those used elsewhere.

By providing evidence about what happens in other systems this report will give a clearer idea of the strengths and weaknesses of methods used in other countries (for clarity we will refer to ‘jurisdictions’ throughout this report as some of the examples studied, e.g. Alberta, are educationally autonomous regions within a country), of what works, what does not work, what might be transferable to England and what is interesting but inappropriate. The evidence provided in this report will assist in reviewing and improving the way standards are maintained as well as helping to explain how standards are set.

This report will discuss the methods used in the context of the educational system within which they operate. This is necessary as they do not operate in isolation but as an integral part of not just the qualifications system but the wider education and social system.

This study is not intended to be an in-depth study of the maintenance of standards and the wider implications in a range of complex systems. It is a brief commentary and a review of the methods used in ten selected jurisdictions. This study will focus on the strengths and weaknesses of each method from a technical perspective – what is done, how it is done and what the issues are. To make meaningful comparisons relevant across a very diverse range of educational systems, this review will focus only on standards in high-stakes, school leaving academic qualifications.

For the purposes of this report, the term ‘maintenance of standards’ relates particularly to achieving comparability of attainment and performance over time and between examinations, for example, between different examination bodies or regions.

1.2 Research aims

There were three aims to this project:

1. To look at how other jurisdictions establish standards when new qualifications are introduced and how they re-set standards if qualifications are reformed.



2. To look at the different methods used to establish and maintain standards across a broad range of jurisdictions.
3. To look at the key strengths and weaknesses of the different methods.

For this study we looked at the systems in Alberta (Canada), Finland, France, Germany, Hong Kong, Massachusetts (USA), the Netherlands, New Zealand, Singapore and South Korea.

1.3 Findings

For the purpose of this review we will use the Ofqual definitions of standards.

- Content standards – what is to be taught and to what level
- Assessment standards – whether the student has learnt what they are supposed to have learned during the course
- Performance standards – how well it has been done

The balance and attention paid to these various forms of standards varies between jurisdictions and tests.

All jurisdictions, in one form or another, used content standards as the foundation for setting standards.

When it comes to setting and maintaining assessment and performance standards, there was no single method that was preferred. New Zealand uses a system relying on expert judgement and, at the other end of the spectrum, Finland and Korea rely on statistical methods. Most jurisdictions rely on a mixed method approach using the strengths of one method to compensate for the weaknesses of another and to allow a wider evidence base to be used to inform decisions when setting standards.

Most jurisdictions studied use some form of expert judgement as part of setting standards, for example having senior examiners looking at sample scripts and determining where the grade boundaries (cut-scores) should be set. The evidence from this review suggests that this approach is important in setting standards as it means the standards are set clearly against what a student can demonstrate they can do but that this approach is not sufficient in itself to allow reliable standard setting. For example, New Zealand, which relies on this approach, has had issues with public perception of standards and other jurisdictions only use this method in combination with other methods, usually statistical, to avoid similar issues.

Most jurisdictions use a variety of statistical methods in determining standards, often to complement expert judgement. Many of these statistical methods use Item Response Theory as it allows common items to be shared between tests and to provide a direct link between the two tests.

Only Hong Kong uses an external reference test to directly set standards. Several jurisdictions have national testing of samples of students to monitor the overall state of the education system and give an overview of standards over time or between regions, often coupled with post-examination reviews or wider research studies.



England appears to be unique in having different assessment authorities (awarding bodies) offering the same level high-stakes national school examinations. There are similarities with the Netherlands, which has a similar problem of equating between examinations provided in different schools, or Canada or Germany where the issue is equating between regions. When it comes to equating between different assessment authorities, e.g. in Germany for the *Abitur*, content standards seem to be the main method. In these systems the regulator is limited in its ability to intervene. Those countries, Germany and Canada, also use national research programmes to monitor standards across the regions. The Netherlands would appear to be a closer comparison to the situation in England and the Dutch regulator uses a wide variety of methods, depending on fitness for purpose, to ensure comparability of standards between tests.

In summary, this research suggests that there is a range of methods in use in the different jurisdictions reviewed and no single method that can deliver all the required aspects of a reliable method for the maintenance of standards. In the majority of cases reviewed, a mixed approach is used, building on the strengths of one approach to compensate for the weaknesses of another. The systems generally feature a combination of statistical methods and expert judgement. This report shows that the balance of methods, and the specific methods used, depends very much on the purposes and context of the examinations.

Many jurisdictions also use some form of external monitoring of their standards either through post-examination research and analysis studies. Only Hong Kong was found to use an external reference test when setting standards each session.



2 Introduction

This report looks at how qualification and assessment standards are set and maintained internationally. That is, how other countries and jurisdictions¹ achieve comparability between student attainment from one year to the next and across different versions of the same qualification (e.g. from different exam boards or different regions). For clarity we will use the term ‘jurisdiction’ to refer to both countries and to sub-regions. For this study we looked at the systems in Alberta (Canada), Finland, France, Germany, Hong Kong, Massachusetts (USA), the Netherlands, New Zealand, Singapore and South Korea. A brief summary of the education systems under discussion is included in Appendix A to help clarify the contexts within which the various methods are being applied.

Maintenance of educational standards is a technically challenging issue and one that attracts a lot of public interest. There has been much debate about the way in which standards are maintained in England. There has also been debate about how educational standards in England compare with those in other countries and whether or not England is “competitive” in the international educational arena. As a result of these concerns, a major educational reform is now underway in England. Part of this reform aims to revise the standards and the grading system for the General Certificate of Secondary Education (GCSE).

Setting standards for a new qualification is a complex technical challenge. In all assessment development it is important to establish the *purpose* of the test and what the results will be used for (Newton, 2007). This allows decisions to be made regarding the results to be reported for the new qualifications/assessments and what these will mean in terms of how the students have performed.

Assessments are used for different purposes in different countries. For example, in many countries students get a high school leaving certificate which shows they have completed a certain amount of schooling and acquired a basic level of attainment in a range of skills. If they want to go to university they will have to sit an additional admissions test. The standards, and the precision of how accurately those standards are measured, are different for these two examinations. In contrast, the GCSE simultaneously serves many purposes; it is a form of school leaving certificate, a means of selection for further education, and is also used for school accountability and measuring national standards. This means it needs to be both an accurate record of the skills of a student and a precise measure of the student’s ability against other students.

For ease of understanding, we need to define what we mean by the term ‘standards’, as different jurisdictions mean different things by the term and have different requirements. To help keep things clear we have mapped these onto the standards as defined by Ofqual (<http://ofqual.gov.uk/standards/>).

¹ A geographical area with separate legal authority and educational system; this could be a country or a state, e.g. Canada which has several jurisdictions, each with a different education system, of which Alberta is one.



- Content standards –what is to be taught and to what level
- Assessment standards – whether the student has learnt what they are supposed to have learned during the course
- Performance standards – how well it has been done

In some jurisdictions performance standards are linked to percentage pass rates – if the students are the expected to be the same then the same percentage will pass.

The balance and attention paid to these various forms of standards varies between countries and tests.

This may seem obvious, but it is necessary to stress these different types of standards as there can be some apparently paradoxical differences between them. For example, it is possible to maintain content and assessment and even the same performance standards, but have widely varying pass rates; or it is possible to maintain the same pass rate but have widely varying performance standards. At times, a mixed-methods approach to standards is used or, within a jurisdiction, the standards used may not be clearly defined. In reality the differences can be less obvious and, as systems evolve due to practicalities and the difficulties of making them work in a robust way, these boundaries can become blurred.



3 Methodology

3.1 Research overview

There were three strands to the research for this study. Firstly, a desk-based review (policy scoping) collected existing information about the education systems of the jurisdictions to be included; secondly, a literature review systematically collected additional information about the jurisdictions being reviewed and the methods they apply for maintaining standards; and finally, a synthesis stage drew together the evidence from the two earlier stages to draw conclusions. More detail about each strand is provided below.

3.2 Policy scoping

Jurisdictions were chosen for the study because they have students who generally either perform well in international surveys or perform at a similar level to English students. They also use a variety of approaches for setting and maintaining standards in high stakes school-leaving academic qualifications. The following jurisdictions were included:

- Alberta – Canada
- Finland
- France
- Germany
- Hong Kong
- Massachusetts – USA
- The Netherlands
- New Zealand
- Singapore
- South Korea.

Summary documents were produced for each jurisdiction and are included as Appendix A, along with a brief overview summary of each jurisdiction.

A rapid desk study – or policy scoping – drawing on existing publicly available evidence sources (INCA² and Eurypedia³) and official websites (education ministries and/or qualifications and assessment authorities) identified the methods used to

² NFER managed and updated the International Review of Curriculum and Assessment Frameworks Internet Archive, known as INCA. It provided descriptions of government policy on education in 21 countries worldwide.

³ Eurypedia is the Eurydice Network's Encyclopedia on National Education Systems, an [online](#) resource for understanding education systems.



establish, maintain and review standards in the selected jurisdictions. The range of jurisdictions included allowed a wide range of methods to be reviewed. For example, New Zealand's National Certificate of Educational Achievement (NCEA) offered the opportunity to look at a criterion-based system, which is very different to the system in South Korea which relies on statistical methods.

A further stage of validating the information was included using international contacts. These contacts were sourced through NFER's work on INCA, and on other international projects including the UK coordination of the PISA, TIMSS and PIRLS international education surveys. This stage was also used to add detail and to clarify any issues which had arisen as the data was collected. Where this additional information is included it is referenced as personal communication (pers. comm.) and where possible supported by independent evidence.

3.3 Literature review

A rapid evidence review was carried out to identify clear and trustworthy evidence about the strengths and weaknesses of the methods used to establish, maintain and monitor standards of high stakes, school leaving academic qualifications in the ten jurisdictions. This involved systematic searching of a range of education and social science bibliographic databases, including the Australian Education Index (AEI), British Education Index (BEI) and ERIC and a consistent best evidence approach to the selection of literature.

The search retrieved 116 records (with some duplication). Of these searches, ten were selected for in-depth review due to their being the most relevant in terms of subject and jurisdictional coverage. In addition to the ten identified by the systematic search it has been necessary to include further records which provide additional clarifications or supporting research within the report. They are listed in the Reference Section of this report along with the selected records.

Brief summaries of each of the ten selected records are provided in Appendix B and fuller details of the process and the search strategy used to locate these records are provided in Appendix C.

3.4 Synthesis

The country profiles produced as a result of the policy scoping and the findings from the literature review were then synthesised and used to provide descriptions of the range of methods used to establish, maintain and review the standards of high stakes school leaving qualifications in each of the selected jurisdictions. The results of this synthesis are provided in the following sections of this report.



4 Findings

4.1 The importance of context

This report looks at the methods used to set and maintain educational standards in ten international jurisdictions⁴: Alberta (Canada), Finland, France, Germany, Hong Kong, Massachusetts (USA), the Netherlands, New Zealand, Singapore and South Korea. It is not a complete technical study of all the possible methods for establishing and maintaining educational standards; nor is it an in-depth study of the education systems in each of the countries listed. Instead, this report gives a picture of a variety of methods that are used to maintain educational standards internationally and some consideration of their strengths and weaknesses. The report discusses the methods in the context of the educational systems in which they are used. This is important because the methods do not operate in isolation but as an integral part of the qualifications system and the wider education and social system in operation in each jurisdiction. It is important to note that most of the methods are not used in isolation but that standards are generally set and maintained by use of a variety of methods and that these need to be discussed together.

4.2 Establishment of standards

Key Findings

New grading systems must be understandable to all users and must signal attainment clearly and reliably.

It is important to have key stakeholder support.

Clearly defined content standards are necessary for the establishment of standards.

Most jurisdictions also use supporting evidence when establishing standards.

When establishing standards it might be necessary to take into account that performance might improve as students get used to a new examination.

Any new grading system must be understandable to all users. It must also clearly and reliably signal the students' attainments so that employers and higher education providers can choose between applicants (Dufaux, 2012; Backes-Gellner and Veen, 2008). If the level of performance needed to gain a particular qualification or grade changes depending on the year that the student took the examination, or who sat the examination, then the results lose their meaning and confidence in them is undermined. This need to set reliable standards can clearly be seen in Germany, where the *Abitur* qualification was losing its perceived value because standards varied between regions, and employers could no longer use it as a reliable signal of

⁴ A note about terminology: the different jurisdictions use different terminology to describe stages of education, examinations and levels of student achievement. When describing the situation in each jurisdiction, we use the terms used by that jurisdiction.



student quality (Backes-Gellner and Veen, 2008). Various programmes are being put in place to rebuild comparability and confidence, for example the introduction of educational standards binding for all Länder and the establishment of the Institute for Educational Progress (*Institut zur Qualitätsentwicklung im Bildungswesen – IQB*) (EACEA, 2013c).

It is also important that any new system has the support of the major stakeholders (schools, parents, employers, higher education providers etc.). The importance of stakeholder engagement, or the problems caused by not involving stakeholders sufficiently, was highlighted during the curriculum reforms that took place in New Zealand in the 1990s (Alison, n.d.; Philips, 2012). Despite a long history of advocacy for standards-based⁵ assessment by their union, many teachers immediately rejected the Government's plans for reform from the old, norm-referenced system to a new standards-based system. The exclusion of teachers from the policy-making process and the political and managerial language used in publications and communications were largely responsible for this rejection (Alison, n.d.).

As described in the introduction, the term 'standards' can be used to mean different things in different situations. Most of the jurisdictions tend to look at 'standards' in two main ways:

- the difficulty and/or the amount of knowledge and skills that students must learn during their course of study - the content standard
- the level of attainment and performance that the student must demonstrate in the examination to be awarded a particular qualification or grade – the assessment and performance standards.

The two are not mutually exclusive and are complementary.

All of the jurisdictions that we looked at, with the exception of the Netherlands, have nationally defined, detailed content standards in the form of 'curriculum frameworks', 'educational standards' or 'programmes of study' that contain descriptions of what students are expected to learn (see Appendix A).

The advantage of nationally defined standards is that teachers and students have a clear idea about what is expected from them (Dufaux, 2012). However, Dufaux highlights that the danger of creating detailed content descriptions linked specifically to a qualification is that too much emphasis may be put on passing the examinations at the cost of other learning, and the curriculum may become limited ("teaching to the test"). This effect can be made worse if some areas cannot easily be tested, because teachers may leave out content that will not be included in the examination (Dufaux, 2012). In the Netherlands, the compulsory national examination syllabuses are short and lack detail, allowing schools to add their own elements to provide tailored education to their students. The final qualification has both a national examination component and a school examination component. It is important that the school curricula are comparable so that qualifications from different schools can be thought of as equivalent. To monitor comparability between school curricula, schools are required to submit their own school examination syllabuses to the Inspectorate

⁵ For an explanation of the differences between these approaches please see 4.3.



showing what elements of the syllabus are tested when, and how marks are calculated, including the weighting of different tests and re-sit opportunities. Non-compulsory guidelines are published to help schools with the task of creating their syllabuses (EACEA, 2013d).

Individual jurisdictions achieved comparability between old and new forms of qualifications in different ways.

In Hong Kong, the performance standards for the new Hong Kong Diploma of Secondary Education (HKDSE) were defined by level descriptors (from Level 1 to 5, with 5** being highest) and illustrated by exemplars of students' work at the various levels of attainment. The performance standards at Levels 4 and 5 of the new HKDSE examination were prepared with reference to the achievements of students who were awarded a grade D or above in the previous Hong Kong Advanced level Examination (HKALE) (HKEAA, 2011). This was necessary to ensure continuity, to aid tertiary selection and to maintain international recognition. When the HKDSE was developed, sample papers were piloted and the results were used as a basis for the development of the level descriptors (HKEAA, 2011).

In Singapore, the standards for the new reduced content 'O' levels and the H1, H2 and H3 examinations were set by expert committees using reference to archive scripts from previous, equivalent qualifications; modelling of expected results; and descriptors of expected performance at key levels. These were developed at the same time as the reformed qualifications were produced (N. Burdett, pers. comm., 2013).

In New Zealand, when the National Certificate of Educational Achievement (NCEA) was first implemented, there was concern among some stakeholders about the varying rigour of some of the standards. When the results of the first (2004) Scholarship examinations were released, there was controversy because the number of Scholarships awarded in some subjects was considerably different from the number in the last year of the old examination. This prompted a public enquiry that made a number of recommendations regarding the administration of the NCEA and resulted in a programme of change that took place between 2006 and 2010 (Philips, 2012).

The Netherlands has not seen the introduction of a significantly new qualification for some time but does introduce new examinations from time to time. When this is the case, there is no single defined method for setting new standards but an ad hoc approach is used to set new qualifications standards, based on the most appropriate method and often using a range of evidence (N. Burdett, pers. comm., 2013). The standards for the central examinations at the end of secondary education are set by the regulator. For new standards, content-based standard setting is often used to ensure the intended standard meets the requirements. Standards are set initially using expert judgement to decide where the grade boundaries should be set (the cut-scores). The most common methods for doing this are extended Angoff or bookmark procedures (Mitzel *et al.*, 2000; Angoff, 1971). In the Angoff method experts make judgements about the probability of a student performing just at the standard getting each question right. This information is then used to estimate what overall mark a



student just achieving the standard will be expected to score. In the bookmark method, the questions are arranged in order of difficulty and experts place a 'bookmark' at the point they think a student just achieving the standard will no longer be able to answer correctly.

For some examinations the authorities in the Netherlands use information from a pre-test which links to an old form of the examination. For others, they use content judgements by experts and, in other cases, they use an assumption of equivalent groups and statistical equating. In this, they also take into account performance on related subjects. For example, foreign languages are considered comparable to each other so, if the performance in a new German language examination has increased, then this is an indication that something similar could be happening in French. Standard setting of new qualifications in the Netherlands also takes into account the assumption that performance in a new qualification will probably increase during the first years (N. Burdett, pers. comm., 2013). This assumption is not unique to the Netherlands and Stobart also refers to this:

Robert Linn (2000) has identified another effect of high-stakes accountability testing movements in America over the last 50 years: scores invariably go up in the first four or so years of a test's implementation and will then plateau. Dan Koretz and colleagues (1991) have also shown that if, after this four-year period of improved scores, students were given the previous test, which the new one replaced, then their scores fell back to where they were in the first year of the new test. So these improvements are essentially about getting better at the test.

(Stobart, 2008)

Whilst taking this into account ensures that overall percentages remain stable and ensures fairness to students in the transition period, it does highlight a debate about how to balance the needs of individuals (in this case altering the standard so that individuals are not disadvantaged by sitting an examination in a certain year) and systemic standards (making sure the standards are equivalent regardless of the impact on individuals).

4.3 Maintenance of standards

Key Findings

The jurisdictions studied employ a wide variety of complementary methods for maintaining standards.

The main forms used are either some version of expert judgement or of statistical techniques or most often a combination of the two.

Only New Zealand relies primarily on expert judgement and this has led to issues of public trust in the standards.

Expert judgement methods might be better at linking results to what student can do but they are limited in their accuracy.

Statistical methods can ensure better year-on-year stability in pass rates but do not, by themselves, reflect content or assessment standards.



In this report, maintenance of standards refers to achieving comparability of attainment and performance over time and between different examination providers, e.g. different school-based examinations or different regions. This is different to the setting of standards for new qualifications described in the previous section.

All of the jurisdictions that we looked at put student results into some form of category: either pass/fail or a grade to show a relative level of achievement. The techniques used to translate the results of the examinations into categories differ between jurisdictions and are often highly complex. For the results to be useful, a 'pass' or a particular grade has to have a clear meaning, and this meaning must not change over time or between different forms of the same qualification.

Broadly speaking, there are two ways in which test results can be translated into categories. Individuals can be put in a category based either on how they performed in comparison to others (known as norm-referencing), or on how their performance compared to a set of pre-determined criteria (known as criterion-referencing).

Performance in a norm-referenced system can either be compared to a carefully chosen sample of the population, which has been used to normalise (calibrate) the test, or to a particular group, for example all the students taking a particular test in a particular year (this is often called cohort-referencing). The advantage of a norm-referenced system is that, if students are competing for a limited number of higher education places or employment opportunities, stakeholders can easily select the highest attaining students. However, norm-referenced grading systems run the risk of being unfair if the students are not all from the same reference group. This may be the case, for example, when students from different cohorts are competing for the same employment opportunity. In addition, because the grade provides no information about actual level of performance, a failing grade does not necessarily prove that the student is low performing, and a high grade does not necessarily mean that the student has reached a minimum level of competence. This is especially true if the reference group is small.

In contrast, criterion-referenced systems provide information about what an individual is capable of, without reference to the capabilities of the rest of the population.

In practice, assessment systems often contain elements of both norm-referencing and criterion-referencing. For example, in New Zealand students receive detailed profiles of their achievement showing their credits and grades for each national standard and how they performed compared with all the other students across the jurisdiction (O'Donnell *et al.*, 2012).

4.3.1 Grading systems that are primarily norm-referenced

Of the jurisdictions that we looked at, only South Korea and Finland use only norm-referencing to assign grades in their school leaving examinations.

In South Korea, the raw scores on the nationally administered College Scholastic Achievement Tests (CSATs) are standardised for each subject, using a linear transformation, and then the distribution of the standardised scores is divided into nine parts (stanines) (KICE, 2013; O'Donnell *et al.*, 2012). Students choose five test



subjects and the results are reported by subject. Some of the subject tests allow topic choice but, from the information obtained for this study, it was not possible to determine whether, and by what method, elective topics are equated before scores are standardised.

Finland also uses a norm-referenced system to grade its upper secondary matriculation examinations. The information for Finland suggests that some variation is allowed in these percentages, ‘the relative shares of grades differ somewhat across the various examinations and examination periods’, but does not explain whether this is to allow for ‘clumping’ of the data (each mark may have a small percentage of students on that mark so matching identical percentages will not be possible), or for calibration of scores once the impact of combinations is completed.

4.3.2 Grading systems that are primarily criterion-referenced

Strictly speaking, a true criterion-referenced system would test whether every individual is (or is not) capable of achieving each and every one of the criteria contained within the curriculum framework. Of the jurisdictions that we considered, the system in New Zealand is closest to this concept in that, for each area of learning, different aspects of skills, knowledge and understanding are assessed separately and each aspect can earn a different number of credits. Consequently, results about each student’s efforts are detailed, giving a profile of their achievements for each national standard (O’Donnell *et al.*, 2012). The risk of this approach is that knowledge ends up being split into disjointed units and, whilst it might be clear that a student has learnt something, it might be harder to show how well they have learnt it.

In practice, in large scale, end of schooling assessment, it is not possible to test every criterion within the curriculum framework. Instead, students are assessed using items (individual questions) or tasks that allow them to demonstrate whether or not they have mastered a particular sample of the content and, from this, their total level of mastery is inferred. The students’ results are reported as levels of achievement based on descriptions of typical performance of students at those levels (known as performance standards).

There are concerns over the ability of experts to accurately make fine distinctions between performance:

Grade-worthiness judgements, generally speaking, are not fine-grained: they are a broad beamed searchlight for identifying standards ... we are utilising examiners’ expertise ineffectively by asking them to make fine distinctions between candidates’ performances in terms of gradeworthiness.

(Baird and Dhillon, 2005)

The key strengths and weaknesses of norm- and criterion-referencing are:



- With norm-referenced methods, where students are competing for limited higher education or employment opportunities, stakeholders can easily select the highest attaining students.
- Norm-referenced grading methods may be unfair if the students are not all from the same reference group.
- A high norm-referenced grade does not necessarily equate to minimum competence. Similarly, a failing grade does not necessarily prove that the student is low performing. This is especially true of small reference groups.
- Criterion-referencing can provide detailed information about what a student knows and is able to do.
- Criterion-referencing can lead to knowledge and skills being broken down into such small pieces that they become disjointed.
- Criterion referencing can lead to wider variations in pass rates year-on-year as expert judgements have a limit to their accuracy in terms of distinguishing between two adjacent marks.

4.3.3 Maintenance of standards over time

Comparability is most straightforward if examinations are carefully designed to provide equivalent tests every year. This is done by using a detailed test specification or blueprint that remains unchanged through the life of a particular programme. The test specification describes how the tests should be put together by giving detailed information about how many marks and what type of questions should be used to test each area of knowledge and skills. Evidence for the use of detailed test specifications was found for Massachusetts, Alberta, the Netherlands, Singapore, Finland, Germany, France, South Korea and Hong Kong. If major changes are made to the syllabus or the test specification these must be taken into account when trying to make the results of the new test comparable to the old one.

Even if two tests are based on the same specification, it is still likely that there will be a difference in the difficulty of the two versions. In addition, the tests are usually taken by different cohorts of students. This means that the distribution of scores (mean and spread of the marks) on the two tests could be affected by both the relative difficulty of the tests and the relative ability of the students. There are various methods for unpicking these factors and trying to make sure that a particular grade represents the same level of performance on both tests. However, it is not an exact science and all of the methods have their strengths and weaknesses. In most cases, the method chosen in the jurisdictions that we looked at strikes a balance between the contrasting needs of the education system in which it is used. The methods are described below.

Methods for ensuring that a grade means the same thing from one examination to the next can be divided into two general groups: those that adjust the students' scores to bring them in line with a previous examination so that pre-defined cut-scores can be applied; and those that use expert judgement and/or statistical



evidence to determine where the cut-scores should be placed on the new distribution.

Statistical techniques for maintaining standards

It is possible to equate different forms of a test taken by different groups of students by including a set of common anchor items (items of known difficulty) in each test. By comparing the performance of the different groups of students on the anchor items, the relative ability of the groups of students can be established. For instance, if there is no significant difference in the average performance of two groups of students on the same set of anchor items it can be inferred that the characteristics of the populations taking the two forms are essentially the same. Thus, any differences in student performance on the unique items in the two forms of the test are the result of differing item difficulties.

The state-wide Massachusetts Comprehensive Assessment System (MCAS) uses item response theory⁶ (IRT) to equate each test to the previous tests. For all Grades (year groups), except Grade 10 (the final year of compulsory education), equating items, which have appeared on previous tests, are used to place the new items on a single difficulty scale. Student raw scores can then be transformed to this scale and fixed cut-scores are applied. At Grade 10, the entire test comprises items that have been previously ‘field-tested’ and, thus, have known difficulties. In this case, no post-test equating is conducted; the student raw scores can be directly transformed to the scale and the fixed cut-scores applied. The cut-scores are fixed every year and are only changed if the standard is reset (MDESE, 2001b).

Alberta also uses anchor items to establish the relative ability of the cohort, and determine whether the current examination is more or less difficult than the baseline examination. A range of equating methods is used, including equipercentile equating (looking at which scores have the same cumulative percentages of students reaching them). Based on the convergence and trends found among these methods, an equating table is developed, which adjusts the students’ total scores to the same standard found in the baseline examination (Alberta Education, 2008). Fixed cut-scores are then applied.

Anchor items and tests can provide a very robust and publicly trusted method of equating. Seeing how students perform on identical questions year-on-year is, on the surface, one of the most obvious ways to ensure standards.

The statistical tools used for IRT are well understood and can produce robust data, provided that the basic assumptions of the model remain unchanged. One of these assumptions is that the construct, which includes aspects of curriculum and how learners will approach answering the question, remains static. In periods of change it will be necessary to challenge these assumptions and make sure that the items still perform as expected.

⁶ A statistical technique which places items on a difficulty scale, using either link items or link students.



The use of anchor items can give a false sense of precision. Anchor item performance can alter with different cohorts and is susceptible to factors such as prior exposure to similar items. There are methods to mitigate against these effects but there is always the scope for drift.

None of the jurisdictions that use this approach use it in isolation, although as an approach for maintaining standards between years, those that do use it place a high reliance on it.

In the Netherlands, raw scores are transformed to a scale that ranges from 1.0 to 10.0, where 1.0 is the lowest and 6.0 is a pass, to correct for possible differences in difficulty of the examination or possible differences in ability of the examination population with respect to a reference examination and population (van Rijn *et al.*, 2008). The estimation of the equating constant relies heavily on IRT data, as well as utilising other statistical methods and expert judgement.

Expert judgements for maintaining standards

In various jurisdictions, including Hong Kong, the Netherlands, New Zealand and Singapore, an expert panel of judges determines the position of the cut-scores so that each grade category represents the same level of achievement each year. The evidence used for these decisions can include: level descriptors, exemplars of students' work, senior examiner recommendations, data from monitoring tests, past levels of performance and other general statistical data that provide information about the relative difficulty of the examinations and the relative ability of the cohort.

In most of the jurisdictions that use this approach, some form of statistical information is also used. This appears to recognise that professional judgement by itself does not provide a reliable enough indication of where to place cut-scores and that statistical data can provide valuable additional information. However, the different sources of information can provide different answers, so expert judgement is required to make sense of the complex and sometimes contradictory evidence. Only New Zealand does not mention use of additional evidence beyond expert judgement when determining cut-scores; the issues of using this approach in New Zealand have been already discussed in section 4.2 (NZQA, 2013e).

The key strength of the professional judgement plus statistical information approach seems to be the expert triangulation of evidence and the ability to combine criterion- and norm-referenced evidence to give a best estimate of where the cut-scores should be. This helps to ensure that no single factor, or individual, can dominate the process. This is particularly important in periods of change where data is likely to be incomplete and different sources may show different results.

The weakness of this approach is that panels of experts are likely to be subject dependent. This may lead to differences in weighting of the evidence between panels, which could influence standards.⁷

⁷ From the information obtained for this report it was unclear how grade boundaries are assigned in the examinations used in France and Germany, for example.



Professional judgement of teachers

Several of the jurisdictions studied use professional judgement of teachers to determine standards. For example, in New Zealand, students gain credit when they demonstrate the knowledge or skills specified in the national standards. Teachers use their expert judgement to decide whether to award the credit and at what level (Achievement, Merit or Excellence). Moderators check a sample of around ten per cent of each school's internally assessed work and adjustments are made to the marks as necessary, although details of how these adjustments are made were not available for this study. Appendix A contains more details about the moderation process and the feedback and teacher training used to ensure that standards are consistently applied.

4.3.4 Comparability between assessment providers or regions

England appears unique in the range of countries studied in having multiple assessment bodies for the main national examinations. However, several of the jurisdictions do have a similar issue, either in equating between assessment bodies across jurisdictions (for example, in Canada, in making sure the standards in Alberta are similar to British Columbia etc., or in Germany across the different *Länder*) or ensuring that the standards applied by different schools or groups of schools are the same (for example, in the Netherlands or New Zealand).

The level to which assessment responsibilities are shared between schools and central agencies varies between the jurisdictions. For the qualifications under consideration in this study:

- In Massachusetts (MCAS), South Korea (CSAT), France (*Baccalauréat* examinations) and Singapore⁸ the student's final grade is composed entirely of externally set and marked elements.
- In Alberta, Hong Kong, the Netherlands, New Zealand, and some *Länder* in Germany, the students' final marks are determined by a combination of public examinations and school-based assessment.
- In Finland and some *Länder* in Germany, students leave compulsory lower secondary education with a leaving certificate based entirely on teacher assessment. Further, in some *Länder*, the upper secondary *Abitur* is also awarded on the basis of school examinations only, however there is a steady move away from decentralised testing in Germany (Dufaux, 2012).

This devolution of assessment responsibility from a single central authority has both strengths and weaknesses. For example, externally processed assessments allow transparency and comparability of student achievement while school-based assessment is believed to enhance validity by including outcomes that cannot easily be assessed within the context of one-off public examinations. In addition, the

⁸ Singapore does use coursework but in many cases this is more similar to controlled assessment and can be considered an external examination, albeit marked by teachers.



results of assessment based on student performance over an extended period of time and developed by those who know the student best can provide a more reliable measure of student ability than a single examination. School-based assessment can help to motivate students by engaging them in meaningful activities, can improve learning by reinforcing curriculum aims and encourage good teaching practice by providing structure and significance to an activity (HKEAA, 2008).

Decentralised assessment (including school-based assessment) has the weakness that it may lead to a lowering of standards. This comes about because employers and higher education institutions cannot adequately distinguish between the different regions or schools, and so individual schools or regions have an incentive to lower their standards and benefit from being pooled with others for whom higher standards have been applied. This undermines the quality of the qualification because it is no longer true that only those of the highest ability possess it, and so qualifications from institutions with higher standards become less meaningful and so they too lower their standards. This “race to the bottom” reduces trust in the certification and employers and higher education institutions can no longer use them as reliable indicators of student ability (Dufaux, 2012).

This effect has been demonstrated in Germany where, although the *Abitur* qualification is recognised across Germany and students can use it to gain access to employment or further education in any state (*Land*), there are, however, large differences between the *Länder* in terms of the percentage of students passing the qualification and the average grade attained (Goethe Institute, 2011). The proportion of students who are *Abitur*-holders compared with non-*Abitur*-holders was significantly lower in *Länder* with central examinations (Backes-Gellner and Veen, 2008; Goethe Institute, 2011). In addition, the national proportion of *Abitur*-holders rose from around ten per cent in 1970 to around 30 per cent in 2002, and this rise was faster in *Länder* without central examinations (Backes-Gellner and Veen, 2008).

Students holding an *Abitur* receive a wage premium compared with their non-*Abitur*-holding peers, however this premium was the same regardless of whether the *Abitur* involved central examinations or not, indicating that employers did not distinguish between the different forms. Further, the wage premium for *Abitur*-holders decreased continually between 1984 and 2002, indicating that employers no longer valued the qualification as highly as they once did. Backes-Gellner and Veen (2008) conclude that:

graduates from states with central examinations and higher educational standards ‘pay’ for the quality deterioration of educational standards in states without central exams.

(p. 569)

It is argued, therefore, that a combination of teacher assessment and common externally set examinations may provide the most accurate picture of the student’s achievement (Alberta Education, 2007). Dufaux (2012) argues that this combination can also prevent the “race to the bottom” while still providing incentives for individual institutions to set higher standards – higher standards have been shown to improve student motivation and performance.



Various methods exist to help to maintain the standards of school-based assessments. For example, Hong Kong provides a detailed description of the guidance, training and moderation measures that are in place to ensure parity in school-based assessment (HKEAA, 2009). Evidence was found for centralised content standards and assessment frameworks for all the jurisdictions that relied on school-based assessments (with the exception of the Netherlands, where schools are encouraged to develop their own curricula). These frameworks detail content to be included, level of achievement expected, assessment methods, subject weighting etc.. In addition, all jurisdictions appear to provide training, guidance and documentation to help schools standardise their examinations. Further details can be found in Appendix A.

Detailed information about the moderation of school-based assessments was found for Alberta, Hong Kong and New Zealand, see Appendix A for details. In Hong Kong, statistical models are used to adjust the spread of the school-based assessment scores of students in a given school with reference to the public examination scores of the same group of students. This process may adjust the students' scores up or down but does not change the rank order assigned by the school. Hong Kong also collects samples of students' work for inspection by professional moderators (HKEAA, 2008). In New Zealand, ten per cent of each school's internally assessed work is checked by professional, full-time moderators (NZQA, 2012b; Philips, 2012). In general, there is good agreement between the marks of the teacher and the moderator. Where agreement is inadequate, scores are adjusted⁹ or the school may be stopped from examining those standards where there is cause for concern (NZQA, 2012b).

In the Netherlands, the external examinations and the school-based assessments contribute equally to a student's final grade in each subject. The Dutch Inspectorate requires that the discrepancy between internal school-based assessment and external national examinations be no more than 0.5 points (on a scale from 1 to 10), and expects schools to analyse and minimise this discrepancy. Research shows that the central examination results are, on average, lower than the school examination results and that the discrepancy is not uniform across schools or student groups (gender and ethnicity were both found to have an effect). These results indicate that the school examinations differ by school and by student group, which may negatively affect the quality of the final certification, leading to a lack of trust in the results (Schildkamp *et al.*, 2012).

In Finland, the national matriculation examinations at the end of upper secondary education (age 19) are marked in the student's school and then sent to the Matriculation Examination Board for centralised marking (SQA, 2008). A similar situation appears to occur in German *Länder* with central *Abitur* examinations. However, in both cases it was unclear whether the centralised marking is an independent double mark (and if so, how the scores are then combined to reach a final mark) or simply a verification of the teacher marks and whether this differs from state to state. In the Netherlands, the examination scripts are marked by the

⁹ Details of these adjustments were not available for inclusion in this report.



students' own teachers and checked by a teacher from another school (Schildkamp *et al.*, 2012).

No information was found for Alberta regarding moderation arrangements or whether teacher-awarded scores are adjusted in any way. This is not to say that there is no moderation of school-based examinations in Alberta, rather that this study did not find any information as to how moderation is conducted. Teacher-awarded marks are compared against the externally-marked examination marks and so it is likely that this information is used to assist teachers in marking to the right standard.

In all the jurisdictions studied, with the exception of Germany, the main high stakes external assessments under consideration are administered by a single body, which is usually the Ministry of Education or another government-appointed organisation. As a result, all students sit the same national examination. In Germany, external *Abitur* examinations are set and marked regionally but they are all set on the same content standards to ensure some comparability.

4.3.5 Comparability between subjects

Subject choice is an important incentive for students because it allows them to focus their efforts on topics of personal interest (Smyth and Hannan, 2007). However, free choice of subjects can lead to problems of comparability. If students are competing for the same higher education or employment opportunities with different combinations of individual subject qualifications (e.g. Singapore 'O' Levels), or with composite qualifications (e.g. high school diploma) made up of different combinations of subjects, those making the selection need some grounds on which to make comparisons. There is no guarantee that the same grade means the same thing across different subjects: the criteria in one subject may be intrinsically easier than another. Similarly, in norm-referenced systems, such as South Korea, if the reference populations are different for different subjects the grades will not be comparable.

It has been demonstrated that subject choices are often based on previous achievement (Lund, 2008) and students prefer subjects that they feel are less rigorous in the expectation that they will do better (Lamprianou, 2009).

Hong Kong, Alberta, Massachusetts, the Netherlands, New Zealand, France and Germany all award certificates based on a combination of final examinations in different subjects. Choice is available in all these systems but, in every case, this choice is always within a restricted framework. The complexity of the framework varies between jurisdictions as does the way in which examination results are combined.

Massachusetts appears to be the simplest system; to be eligible for the high school diploma students must (among other things) pass MCAS examinations in English language arts and mathematics and in one of biology, chemistry, introductory physics or technology/engineering. A decision rule of this form has the advantage that it ensures that students meet minimal standards in key subjects. However, the simple rule "must pass all" has been shown to lead to more misclassification (i.e. students



receiving the wrong decision) compared with decision rules that allow compensation between subjects. That is, systems that allow good performance in one subject to make up for poor performance in another. The system for awarding the final diploma in the Netherlands is an example of this sort of compensatory decision rule (van Rijn *et al.*, 2008).

In France, the *Baccalauréat* is awarded if students obtain an average score of 10 out of a possible 20 in their final examinations. However, systems of this sort, that combine results using a simple averaging process, can lead to a systematic bias against students enrolled in more rigorous subjects.

In some jurisdictions, such as in South Korea, standardised scores are reported; raw scores are converted onto a scale with a pre-defined mean and standard deviation. However, if the original population is not uniform, is highly skewed, or has an extreme mean or standard deviation, such conversions will not produce comparability between subjects.

Various other statistical models, including item response theory (IRT), can be used to adjust raw scores to account for the relative difficulty of the courses or examinations and the differences in the ability of students (Lamprianou, 2009). These models can be elaborated to remove the assumption that proficiency is uni-dimensional (i.e. that the trait that makes an individual good at mathematics is the same as the one that makes them good at English), and also to take into account the assumption that more proficient students may choose subjects that are more difficult (Korobko *et al.*, 2008).

The disadvantage with these techniques is that they very quickly become very complicated and produce reports that are unintelligible to most audiences, leading to mistrust in the system. In addition, the results are meaningless if the data does not conform to the assumptions of the model, for example, if the population of students taking an examination is not uniform. A further consequence of scaling scores is that it can affect student choice, with students avoiding subjects that are routinely scaled down (Lamprianou, 2009).

The German *Abitur* is awarded on the basis of a combination of the results of all subjects taken, but no details were found as to the method used to calculate the final grade.

New Zealand (NCEA) and Alberta (High School Diploma) allow students a great deal of flexibility to choose which courses they follow to gain credits towards their final certificate (see Appendix A for details). No information could be found as to how Alberta accounts for differences in the ease of gaining credits in some courses compared with others, and minimal information was available for New Zealand.

When the NCEA was first implemented, the proportion of results awarded at each of the three levels (Achievement, Merit and Excellence) varied from standard to standard and within a particular subject. Such variation was not considered problematic by central authorities – it was simply accepted that some standards were harder to achieve than others. However, schools, teachers and parents were concerned with the variability. In addition, students appeared to be adopting



strategic approaches to collecting credits, on the basis of those that were easier to obtain and those which would allow demonstration of higher levels of achievement. To improve comparability, standards were more tightly defined where necessary and levels of achievement were closely monitored. A scaling system was implemented for results that were outside bands of acceptable tolerance (Philips, 2012).

It appears that comparability per se may not be as important as the perceived fairness of the system (Lamprianou, 2009) and it could become unimportant if society can accept that subjects vary in their rigour and use the results accordingly. Information on the relative performance of students in each subject compared with the performance of the same students across other subjects could be reported for stakeholders to use as they see fit.

The alternative is to limit subject choice in such a way that stakeholders are only ever comparing like with like (as is the case in Greece, Lamprianou, 2009).

4.4 Reference testing

Key Findings

Only Hong Kong use reference tests for setting standards in each session.

Evidence for the use of reference tests was found only for Hong Kong, where monitoring tests were administered under secure conditions in the core subjects to the last cohorts of students to take the old examinations in 2010 and 2011 respectively. The same tests were administered to a sample of students taking the new examinations in 2012. The statistical data from the monitoring tests was used to provide linkage between the standards of the new examinations and the previous examinations, as well as standards going forward in time.

The information from South Korea suggests that results in the CSAT are “linked to data of another similar national-level test...” but it was not possible to verify this or confirm how this was done (KICE, 2012).

4.5 Post-examination reviews

Key Findings

Many jurisdictions conduct post-examination statistical evaluations.

Many jurisdictions seek feedback, pre- or post-examination, from a range of stakeholders about the appropriateness of the examination.

Some of the jurisdictions studied also have mechanisms for monitoring national educational standards over time and for feeding back to sub-jurisdictions.

Evidence was found for psychometric and statistical scrutiny of examination results and test functioning in Massachusetts, Alberta, Hong Kong, Finland and the Netherlands. Massachusetts and Hong Kong both publish full technical reports, see Appendix A for details. CITO, the Dutch National Institute for Educational Measurement, publishes psychometric analyses for the national tests conducted in



the Netherlands, but little information is available about the school-based assessments (van Rijn et al., 2012). In Finland, there is statistical research and analysis of the results of the matriculation examinations by test and by censor (marker) (Finnish Eurydice Unit, 2013), but details of these analyses and information as to whether the results are published was not available. Similarly, Alberta conducts post-examination statistical analyses for the purposes of equating tests, but it is unclear as to whether this information is published.

Finland, Hong Kong and the Netherlands all seek feedback from various stakeholders (including schools and universities) regarding the latest session of examinations and/or conduct system audits. The results of the feedback/audits are used to improve the examinations and the surrounding processes. In Hong Kong, examination papers and student responses are also sent overseas for external review. The aim of this process is to collect feedback on the paper and on marking standards, which can be used to inform improvement, and to ensure that standards are comparable with those of overseas examinations.

In Alberta some pre-examination review occurs to ensure that a panel of selected teachers feel that the examination questions are appropriate.

This study did not find information on formal, regular reference testing or post-examination review systems in Singapore, Alberta, France, New Zealand (other than at primary level) or South Korea (see note in Section 4.4 above), although this does not necessarily mean that such measures do not exist. In Canada and Germany, which are federal systems, standards are also monitored over time and across jurisdictions, and research, monitoring tests (e.g. the Pan Canadian Assessment Programme) and the results of international tests such as PISA feed into the maintenance of standards.



5 Conclusion

The clear finding from this research is that most jurisdictions use a mixed approach, supplementing the weaknesses of one method with the strengths of another and looking at a broad range of evidence to inform the setting and maintenance of standards.

All jurisdictions use some form of content standards, defining the minimum curriculum to be covered, as the basis for ensuring comparability between examinations and over time. For the Netherlands, these are the least detailed but the Inspectorate ensures comparability of curricula offered.

The jurisdictions studied employ a wide variety of approaches to setting and maintaining standards. The main forms used are either some version of expert judgement or of statistical techniques, with the simplest being norm-or cohort-referencing. The use of pure assessment standards and expert judgement (including teacher judgement) of standards can lead to unacceptable variation in numbers of students achieving key boundaries which can cause issues of public mistrust. Only New Zealand relies primarily on this method and has had a long battle to make it work to an acceptable level.

Two countries (Finland and Korea) use the percentage of students passing each year as a measure of performance standards (i.e. norm-referencing). This works well and seems to be trusted and stable in those countries, but does mean there is no direct link to what students can do and demonstrate and so, as a method of ensuring standards, it could be limited.

Several use methods which allow direct equating of a test via anchor items or reference tests to previous years' standards, often using Item Response Theory (IRT).

Most jurisdictions also have some means of pre- or post-examination review of standards or are in the process of establishing them. This includes the HKEAA in Hong Kong sending scripts internationally for them to be reviewed by experts. Some countries also use reference to tests such as PISA to monitor their standards over time.



References

- Alberta Education (2007). *Meeting High Standards* [online]. Available: <http://www.education.alberta.ca/parents/educationsys/ourstudents/viii.aspx> [4 October, 2013].
- Alberta Education (2008). *Maintaining Consistent Standards Over Time Initiative* [online]. Available: <http://education.alberta.ca/admin/testing/diplomaexams/standards-.aspx> [4 October, 2013].
- Alberta Education (2013). *Guide to Education: ECS to Grade 12 2013-2014* [online]. Available: <http://www.education.alberta.ca/media/7044993/guidetoed2013.pdf> [4 October, 2013].
- Alison, J. (n.d.). *A Lesson in How not to Make Education Policy Change: New Zealand School Qualifications and the Struggle for Hegemony* [online]. Available: http://www.acsa.edu.au/pages/images/judie_alison.pdf [4 October, 2013].
- Angoff, W.H. (1971). 'Scales, norms and equivalent scores.' In: Thorndike, R.L. (Ed) *Educational Measurements*. Washington, DC: American Council on Education.
- Backes-Gellner, U. and Veen, S. (2008). 'The consequences of central examinations on educational quality standards and labour market outcomes', *Oxford Review of Education*, **34**, 5, 569–588.
- Baird, J-A. and Dhillon, D. (2005). *Qualitative Expert Judgments on Examination Standards: Valid, but Inexact* (AQA Research Report RPA_05_JB_RP_077). Guildford: AQA.
- Card, J. (2005). 'Life and death exams in South Korea', *Asia Times*, 30 November [online]. Available: <http://www.atimes.com/atimes/Korea/GK30Dq01.html> [4 October, 2013].
- Center on Education Policy (2012). *MASSACHUSETTS Profile of State High School Exit Exam Policies 2012*. Washington, DC: Center on Education Policy [online]. Available: <http://www.cep-dc.org/documents/HSEE2012Profiles/MassachusettsHSEE2012.pdf> [4 October, 2013].
- College voor Examens (2013). *College voor Examens* [online]. Available: <http://www.cve.nl/> [4 October, 2013].
- College voor Examens and CITO (2011). *Toets Special: Normering* [online]. Available: <http://toetswijzer.kennisnet.nl/html/normering/default.shtm> [4 October, 2013].
- Commonwealth of Massachusetts (2013). *General Laws* [online]. Available: <https://malegislature.gov/Laws/GeneralLaws/PartI/TitleXII/Chapter69/Section1D22/09/13> [4 October, 2013].



Coniam, D. (2011). 'The double marking of liberal studies in the Hong Kong public examination system', *New Horizons in Education*, **59**, 2, 1–12.

Dufaux, S. (2012). *Assessment for Qualification and Certification in Upper Secondary Education: a Review of Country Practices and Research Evidence* (OECD Education Working Papers, No. 83). Paris: OECD Publishing [online]. Available:

<http://www.oecd-ilibrary.org/docserver/download/5k92zp1cshvb.pdf?expires=1380886666&id=id&accname=guest&checksum=33F3F2F70774600332307DE5D206E96B> [4 October, 2013].

Education Commission of the States (2010). *Compulsory School Age Requirements* [online]. Available: <http://www.ncsl.org/documents/educ/ECSCCompulsoryAge.pdf> [4 October, 2013].

Education, Audiovisual and Culture Executive Agency (2013a). *Eurypedia: Finland* [online]. Available:

<https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Finland:Overview> [4 October, 2013].

Education, Audiovisual and Culture Executive Agency (2013b). *Eurypedia: France* [online]. Available:

<https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/France:Overview> [4 October, 2013].

Education, Audiovisual and Culture Executive Agency (2013c). *Eurypedia: Germany* [online]. Available:

<https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Germany:Overview> [4 October, 2013].

Education, Audiovisual and Culture Executive Agency (2013d). *Eurypedia: the Netherlands* [online]. Available:

<https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Netherlands:Overview> [4 October, 2013].

Finnish Eurydice Unit (2013). Response to EURYDICE Network Question Standards in the Matriculation Examination 19 September 2013. Unpublished report.

Finnish Matriculation Board (2007). *Finnish Matriculation Examination* [online]. Available: <http://www.ylioppilastutkinto.fi/en/> [4 October, 2013].

Finnish National Board of Education (2009). *National Core Curriculum for Basic Education 2004* [online]. Available:

http://www.oph.fi/english/publications/2009/national_core_curricula_for_basic_education [4 October, 2013].

Finnish National Board of Education (2011). *General Upper Secondary Education* [online]. Available:

http://www.oph.fi/english/education/general_upper_secondary_education [4 October, 2013].



Goethe Institute (2011). *Centralizing the Abitur – Equal Conditions for All* [online]. Available: <http://www.goethe.de/ins/cz/prj/zuk/dos/org/en7956574.htm> [4 October, 2013].

Government of The Netherlands (2012) *Final Secondary Education* [online]. Available: <http://www.rijksoverheid.nl/onderwerpen/voortgezet-onderwijs/eindexamen-voortgezet-onderwijs> [4 October, 2013].

Hong Kong Examination and Assessment Authority (2008). *Hong Kong Diploma of Secondary Education* [online]. Available: http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/HKDSE_pamphlet_Eng_1410.pdf [4 October, 2013].

Hong Kong Examination and Assessment Authority (2008). *School Based Assessment (SBA) for HKDSE* [online]. Available: http://www.hkeaa.edu.hk/en/sba/sba_hkdse [4 October, 2013].

Hong Kong Examination and Assessment Authority (2009). *Information on School Based Assessment* [online]. Available: http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/SBA_pamphlet_E_web.pdf [4 October, 2013].

Hong Kong Examinations and Assessment Authority (2011). *Hong Diploma of Secondary Education Examinations: Quality Assurance Framework* [online]. Available: http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/HKDSE_QA_A4booklet_Jun2011.pdf [4 October, 2013].

Hong Kong Examination and Assessment Authority (2013a). *Important Dates of 2014 HKDSE Examination* [online]. Available: http://www.hkeaa.edu.hk/en/hkdse/Event_timetable/ [4 October, 2013].

Hong Kong Examination and Assessment Authority (2013b). *HKDSE* [online]. Available: <http://www.hkeaa.edu.hk/en/hkdse/> [4 October, 2013].

Institut zur Qualitätsentwicklung im Bildungswesen (2012). *Aufgabenpool für die Abiturprüfung* [online]. Available: <http://www.iqb.hu-berlin.de/bista/abi/aufgabenpool> [4 October, 2013].

International Bureau of Education (2010a). *World Data on Education: Canada 2011/12* [online]. Available: <http://www.ibe.unesco.org/en/services/online-materials/world-data-on-education/seventh-edition-2010-11.html> [4 October, 2013].

International Bureau of Education (2010b). *World Data on Education: Finland 2011/12* [online]. Available: [http://www.\(IBE,2010\).unesco.org/en/services/online-materials/world-data-on-education/seventh-edition-2010-11.html](http://www.(IBE,2010).unesco.org/en/services/online-materials/world-data-on-education/seventh-edition-2010-11.html) [4 October, 2013].

International Bureau of Education (2010c). *World Data on Education: Hong Kong 2011/12* [online]. Available: <http://www.ibe.unesco.org/en/services/online-materials/world-data-on-education/seventh-edition-2010-11.html> [4 October, 2013].

ITS (2013). *The Hong Kong Education and Schooling System Explained* [online]. Available: <http://www.itseducation.asia/education-system.htm> [4 October, 2013].



Kaftandjieva, F. and Takala, S. (2002). 'Relating the Finnish Matriculation Examination English test results to the CEF Scales.' Paper presented at Helsinki Seminar on Linking Language Examinations to Common European Framework of Reference for Languages: Learning, Teaching and Assessment. Cited in: Scottish Qualifications Agency (2008). *The Assessment Systems of Finland and Queensland* (Policy and New Products Research Report 1). [online]. Available: http://www.sqa.org.uk/files_ccc/PNP_ResearchReport1_TheAssessmentSystemsFinlandQueensland.pdf [4 October, 2013].

Korea Institute for Curriculum and Evaluation (2000). *A Study on Constructing (Developing) Item Bank for CSAT* [online] Available: <http://kice.re.kr/en/board.do?boardConfigNo=139&page=69&menuNo=410&action=view&boardNo=28419> [4 October, 2013].

Korea Institute for Curriculum and Evaluation (2012). *Education in Korea* [online] Available: http://www.kice.re.kr/design/images/pdf/2013_ek_full.pdf [4 October, 2013].

Korea Institute for Curriculum and Evaluation (2013). *College Scholastic Aptitude Test* [online] Available: <http://www.kice.re.kr/en/contents.do?contentsNo=149&menuNo=405> [4 October, 2013].

Koretz, D., Linn, R., Dunbar, S. and Shepard, L. (1991). 'The effects of high stakes testing on achievement: preliminary findings about generalization across tests.' Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, 5 April [online]. Available: <http://nepc.colorado.edu/files/HighStakesTesting.pdf> [2 December, 2013].

Korobko, O.B., Glas, C.A.W., Bosker, R.J. and Luyten, J.W. (2008). 'Comparing the difficulty of examination subjects with item response theory', *Journal of Educational Measurement* **45**, 2, 139–157.

Lamprianou, I. (2009). 'Comparability of examination standards between subjects: an international perspective', *Oxford Review of Education*, **35**, 2, 205–226.

Linn, R.L. (2000). 'Assessment and accountability', *Educational Researcher*, **29**, 2, 4–16.

Lund, S. (2008). 'Choice paths in the Swedish Upper Secondary Education: a critical discourse analysis, of recent reforms', *Journal of Education Policy*, **23**, 6, 633–648. Cited in: Dufaux, S. (2012). *Assessment for Qualification and Certification in Upper Secondary Education: a Review of Country Practices and Research Evidence* (OECD Education Working Papers, No. 83). Paris: OECD Publishing [online]. Available: <http://www.oecd-ilibrary.org/docserver/download/5k92zp1cshvb.pdf?expires=1380886666&id=id&accname=guest&checksum=33F3F2F70774600332307DE5D206E96B> [4 October, 2013].



Massachusetts Department of Elementary and Secondary Education (1995). *Final Regulations on Evaluation of Educators* [online]. Available: <http://www.doe.mass.edu/lawsregs/603cmr35.html?section=all> [4 October, 2013].

Massachusetts Department of Elementary and Secondary Education (2011a). *Massachusetts Curriculum Frameworks: Current Curriculum Frameworks* [online]. Available: <http://www.doe.mass.edu/frameworks/current.html> [4 October, 2013].

Massachusetts Department of Elementary and Secondary Education (2011b). *2011 MCAS and MCAS-Alt Technical Report* [online]. Available: http://www.mcasservicecenter.com/documents/MA/Technical%20Report/2011/2011%20MCAS%20and%20MCAS%20Alt%20Tech%20Report_Final.pdf [4 October, 2013].

Massachusetts Department of Elementary and Secondary Education (2013). *Massachusetts Comprehensive Assessment System Overview* [online]. Available: <http://www.doe.mass.edu/mcas/overview.html?faq=4> [4 October, 2013].

Ministry of Education, Finland (2007). *Education and Research 2007-2012* [online]. Available: <http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2008/liitteet/opm11.pdf> [4 October, 2013].

Ministry of Education, France (2012). *Baccalauréat 2012* [online]. Available: http://www.education.gouv.fr/cid60617/baccalaureat-2012.html#Baccalauréat_général_et_technologique [4 October, 2013].

Ministry of Education, Singapore (2007a). *The 'A' Level Experience* [online]. Available: <http://www.moe.gov.sg/cpdd/alevel2006/experience/exp.htm> [4 October, 2013].

Ministry of Education, Singapore (2007b). *The 'A' Level Experience – Frequently Asked Questions* [online]. Available: <http://www.moe.gov.sg/cpdd/alevel2006/faqs.htm#Q21> [4 October, 2013].

Ministry of Education, Singapore (2012). *Secondary School Education* [online]. Available: <http://www.moe.gov.sg/education/secondary/files/secondary-school-education-booklet.pdf> [4 October, 2013].

Ministry of Education, Singapore (2013). *Changes Affecting Special/Express Courses* [online]. Available: <http://www.moe.gov.sg/education/secondary/express/changes/> [4 October, 2013].

Mitzel, H. C., Lewis, D. M., Patz, R. J., and Green, D. R. (2000). 'The bookmark procedure: cognitive perspectives on standard setting.' In: Cizek, G.J. (Ed) *Setting Performance Standards: Concepts, Methods, and Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.

National Foundation for Education Research (2012). *Review of the National Curriculum in England: Report on Subject Breadth in International Jurisdictions* (DFE Research Report 178A). London: DFE [online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/197636/DFE-RR178a.pdf [4 October, 2012].



Newton, P. E. (2007). 'Clarifying the purposes of educational assessment', *Assessment in Education*, **14**, 2, 149–170.

New Zealand, Ministry of Education (1997). *Education 1997-1999: Government Strategy. Information Document*. Wellington: Ministry of Education. Cited in O'Donnell, S., Byrne, A., Goram, R., Higginson, C., Sargent, C. and White, E. (2012). *INCA (The International Review of Curriculum and Assessment Frameworks) Internet Archive*

New Zealand Qualifications Authority (2012a). *Changes to Standards and National Qualifications* [online]. Available: <http://www.nzqa.govt.nz/qualifications-standards/changes-to-standards-and-national-qualifications/> [4 October, 2013].

New Zealand Qualifications Authority (2012b). *Factsheet 10: How does NZQA Make Sure Internal Assessment is Fair and Consistent Across the Country?* [online]. Available: <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/the-facts/factsheet-10/> [4 October, 2013].

New Zealand Qualifications Authority (2013a). *Awards* [online]. Available: <http://www.nzqa.govt.nz/qualifications-standards/awards/> [4 October, 2013].

New Zealand Qualifications Authority (2013b). *History of the NCEA* [online]. Available: <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/history-of-ncea/> [4 October, 2013].

New Zealand Qualifications Authority (2013c). *Standards* [online]. Available: <http://www.nzqa.govt.nz/qualifications-standards/standards/> [4 October, 2013].

New Zealand Qualifications Authority (2013d). *Scholarship* [online]. Available: <http://www.nzqa.govt.nz/qualifications-standards/awards/scholarship/> [4 October, 2013].

New Zealand Qualifications Authority (2013e) *NCEA External Assessment: Grade Score Marking* [online]. Available: <http://nzqa.govt.nz/qualifications-standards/qualifications/ncea/ncea-exams-and-portfolios/external/grade-score-marking> [2 December, 2013].

O'Donnell, S., Byrne, A., Goram, R., Higginson, C., Sargent, C. and White, E. (2012). *INCA (The International Review of Curriculum and Assessment Frameworks) Internet Archive*

Philips, D. (2012). 'The impact of research on the evolution of secondary qualifications in New Zealand.' Paper presented at the British Educational Research Association Annual Conference, University of Manchester, 4–6 September.

Schildkamp, K., Rekers-Mombarg, L.T.M. and Harms, T.J. (2012). 'Student group differences in examination results and utilization for policy and school development', *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, **23**, 2, 229–255.



Schoolexamensvo.nl (2012). *Toetsingskader van Cito voor de kwaliteitsbeoordeling van schoolexamens* [online]. Available: http://www.schoolexamensvo.nl/fileadmin/contentelementen/kennisnet/Schoolexamensvo/Scheikunde/Scheikunde_aug_2012/3_Toetsingkader_Cito_.pdf [4 October, 2013].

Scottish Qualifications Agency (2008). *The Assessment Systems of Finland and Queensland*. Available online: http://www.sqa.org.uk/files_ccc/PNP_ResearchReport1_TheAssessmentSystemsFinlandandQueensland.pdf [4 October, 2013].

Singapore Examinations and Assessment Board (2013). *Instructions for Private Candidates Taking the Singapore-Cambridge GCE N(T)-, N(A)-, O- and/or A-Level Examinations* [online]. Available: <http://www.seab.gov.sg/privateExamInstructions/2013InstructionsForPrivateCandidates.pdf> [4 October, 2013].

Smyth, E. and Hannan, C. (2007). 'School processes and the transition to higher education', *Oxford Review of Education*, **33**, 2, 175–194. Cited in: Dufaux, S. (2012). *Assessment for Qualification and Certification in Upper Secondary Education: a Review of Country Practices and Research Evidence* (OECD Education Working Papers, No. 83). Paris: OECD Publishing [online]. Available: <http://www.oecd-ilibrary.org/docserver/download/5k92zp1cshvb.pdf?expires=1380886666&id=id&accname=guest&checksum=33F3F2F70774600332307DE5D206E96B> [4 October, 2013].

Standing Conference of the Ministers of Education and Cultural Affairs of the Länder (KMK) (2008). *Vereinbarung über Einheitliche Prüfungsanforderungen in der Abiturprüfung* [online]. Available: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2008/2008_10_24-VB-EPA.pdf [4 October, 2013].

Standing Conference of the Ministers of Education and Cultural Affairs of the Länder (KMK) (2012). *Creating the Basis for Comparable Abitur Examinations in Germany* [online]. Available: <http://www.kmk.org/information-in-english/press-releases.html> [4 October, 2013].

Standing Conference of the Ministers of Education and Cultural Affairs of the Länder (KMK) (2013). *Übersicht über die Abiturtermine 2014* [online]. Available: http://www.kmk.org/fileadmin/pdf/Bildung/AllgBildung/Abiturtermine_2014.pdf [4 October, 2013].

Stobart, G. (2008) *Testing Times. The Uses and Abuses of Assessment*. London: Routledge.

University of Massachusetts Amherst (2011). *Psychometric Methods, Educational Statistics, and Research Methods (Psychometrics)* [online]. Available: http://www.umass.edu/remf/news_MCASreports.html [4 October, 2013].



VO Raad (n.d.). *Standaard Borging Kwaliteit Schoolexamens* [online]. Available: http://www.vo-raad.nl/userfiles/bestanden/VO-raad_Standaard-schoolexamens.def.pdf [4 October, 2013].

van Rijn, P.W., Béguin, A.A. and Verstralen, H.H.F.M. (2012). 'Educational measurement issues and implications of high stakes decision making in final examinations in secondary education in the Netherlands', *Assessment in Education: Principles, Policy & Practice*, **19**, 1, 117–136.



Appendix A – Summaries of jurisdiction education systems

This Appendix contains tables which provide an overview of the education system and of the main high stakes school leaving examinations in the ten jurisdictions of this study. The information relates to seven research questions below. Where a question has been left blank, no information was available.

- Period of compulsory education?
- What high stakes school leaving academic qualifications exist and when are they taken?
- What format do they take, how does testing take place and what is tested?
- What methods are used to establish standards (for each relevant qualification)?
- What methods are used to maintain standards over time (for each relevant qualification)?
- Do monitoring assessments exist?
- What methods are used to review standards, e.g. are there post-examination reviews, and how are the results used?

Before the tables commence, we provide a one paragraph summary of the key features of the high stakes testing system in each jurisdiction.

Alberta (Canada) – compulsory education lasts for ten years from age six/seven to age 16, but most students continue to around age 18. Schooling is split into primary and intermediate at the elementary level, and compulsory junior high school and post-compulsory senior high school at the secondary level. Students do not obtain a formal leaving qualification at the end of junior high school (Grade 9, age 15), but provincially-administered achievement tests are used to determine the most suitable form of senior high school for an individual student. Various qualifications are available at senior high school, with the majority of students receiving the (Alberta) High School Diploma (NFER, 2012).

Finland – compulsory ‘basic’ education lasts for nine years, from age seven until age 16. Post-compulsory, upper secondary education is available to 16-19 year olds in general and vocational upper secondary schools. There are no national examinations in basic education; all assessment is in-school and completed by teachers. All students who complete the full nine years of compulsory education and receive a grade of ‘adequate’ in their in-school assessment are awarded the basic education certificate (FNBE, 2009). Selection for upper secondary school is based on students’ grade point average for the academic subjects in the basic education certificate (EACEA, 2013a). A national matriculation examination is taken at the end of upper secondary education (at age 19), the results of which contribute to the eligibility



requirements for entry to higher education (EACEA, 2013a; IBE, 2010b). The grading system for these examinations has six categories of pass, from the lowest A (*approbatur* = passed) to the highest L (*laudatur* = praised), and I (*improbatur* = failed) (EACEA, 2013a).

France – compulsory education commences at age six and ends at 16. The major high stakes academic school leaving qualification is the general *Baccalauréat*. The general *Baccalauréat* examinations are administered nationally and reflect the courses offered in the final two years of general education (ages 16-18). Successful completion entitles the holder to enter higher education. (O'Donnell *et al.*, 2012).

The Federal Republic of Germany – compulsory education lasts from age six to 18/19, the detailed organisation of which differs between *Länder* (the administrative states into which Germany is divided). On successful completion of lower secondary education, at age 15/16, students are awarded one of a range of leaving certificates by their school. In an increasing number of *Länder*, students are required to sit a central examination at state level to obtain this leaving certificate. Upper secondary education concludes, at the age of 18/19, with the *Abitur* examination. Successful completion of the *Abitur* results in the award of the *Allgemeine Hochschulreife* which is required to enter higher education. Students are awarded a grade from 1 (*sehr gut*, very good) to 6 (*ungenügend*, unsatisfactory) (O'Donnell *et al.*, 2012). The exact timing and subject choices for the *Abitur* examinations vary between individual states (EACEA, 2013c).

Hong Kong – education is compulsory for six years of primary schooling and three of lower secondary, age six to age 15 (IBE, 2010c). Students in post-compulsory upper secondary education (age 15-17) prepare for the Hong Kong Diploma of Secondary Education (HKDSE) (ITS, 2013). The grade set for HKDSE runs from 1 (lowest performance) to 5, with 5** available for the highest performances. Performance below Level 1 is labelled as unclassified (HKEAA, 2013b).

Massachusetts (USA) – education is compulsory for ten years between the ages of six and 16 (ECS, 2010). The Massachusetts Comprehensive Assessment System (MCAS) aims to test all public school students across the state, including students with disabilities and those with limited English proficiency. The MCAS measures student performance based on the Massachusetts curriculum framework learning standards. Passing the MCAS Grade 10 tests (end of compulsory education) is a requirement for eligibility for a high school diploma (awarded on completion of Grade 12, aged 18) (MDESE, 2013). The MCAS reports at four levels: advanced, proficient, needs improvement and warning/failing (O'Donnell *et al.*, 2012).

The Netherlands – education is compulsory from age five to age 16, although all young people under the age of 18 must be in some form of education until they achieve a basic qualification: the HAVO, VWO or VMBO. VWO is the major high stakes school leaving qualification, which provides access to university, higher professional education and employment. A scale of 1-10 is used for the final examinations, with 1 being the lowest and 6 constituting a pass (EACEA, 2013d).

New Zealand – education is compulsory from age six to 16. Secondary education caters for students aged 13 to 18 years. The National Certificate of Educational



Achievement (NCEA) is the main qualification at all levels of senior secondary school. Students can achieve the NCEA from a wide range of studies, within and beyond the school curriculum. Achievement standards (for general/academic school subjects) and unit standards (for vocational/technical subjects) specify what a person knows and can do in a particular area of learning. Achievement standards are awarded at three levels: achievement, merit and excellence; unit standards are passed at one level only. Credits are gained when students attain these standards, and an NCEA is awarded when the student gains a specified number of credits at a particular level of the National Qualifications Framework. NCEA (and other national certificates) are recognised by employers and used as the benchmark for selection by universities and polytechnics. (O'Donnell *et al.*, 2012).

Singapore – students generally complete ten years of formal education, from age six to 16/17. There are six years of compulsory primary school education and, although not compulsory, most students continue for at least another four years of secondary education. Post-secondary, or pre-university, education comprises a further two to three years for students aged 16/17 to 18/19/20. There is a series of high stakes examinations, of which the Singapore-Cambridge General Certificate of Education Normal Level (GCE 'N' Level) and GCE Ordinary Level (GCE 'O' Level) examinations are of most relevance to this report. The GCE 'N' Level or GCE 'O' Level are taken at the end of secondary education (age 16/17) and provide access to post-secondary academic education, or various technical or business study courses (O'Donnell *et al.*, 2012). The grading system for the GCE 'N' and 'O' levels is 1 to 5 or U (ungraded).'

South Korea – education is compulsory for nine years between age six and age 15 (O'Donnell *et al.*, 2012). Post-compulsory upper secondary (high school) education (offering advanced general and specific education) is provided to students aged 15-18. On completion of high school education, successful students are awarded a diploma. However, it is the nationally administered College Scholastic Ability Tests (CSATs) that are most important for access to higher education (KICE, 2012). Results are reported on a nine-point scale, with 1 being the highest.



Alberta, Canada

Period of compulsory education

Compulsory education in Alberta lasts for ten years from age six/seven to age 16 but most students continue until around age 18 (Year/Grade 12). The public education system is divided into two levels, elementary and secondary. Each, in turn, is subdivided into two components: primary and intermediate at the elementary level and, at secondary level, compulsory lower secondary education in junior high school (Grades/Years 7 to 9 – students aged 12/13 to 14/15) and post-compulsory upper secondary education in senior high school (Grades 10 to 12 - students aged 15/16 to 18 years) (O'Donnell *et al.*, 2012).

What high stakes school leaving academic qualifications exist and when are they taken?

There is no formal leaving qualification at the end of lower secondary phase education, Year/Grade 9, age 15. However, provincial achievement tests are administered annually on completion of lower secondary education. The subjects tested in Year 9 are mathematics, science, social studies, English and French (O'Donnell *et al.*, 2012).

The results of continuous classroom assessment or of the province-wide tests do not normally influence a student's possibilities of promotion to the next class. However, the results of the Grade/Year 9 province-wide tests, taken on completion of junior high school, aged 15, are often used, in combination with advice from guidance counsellors, to determine the type of senior high school most suited to a student's needs. The results of these tests are generally regarded as being good predictors of high school performance (O'Donnell *et al.*, 2012).

The following qualifications are available at the end of upper secondary education (age 18):

- Alberta High School Diploma (English and Francophone)
- Certificate of High School Achievement (English and Francophone) (for certain students enrolled in knowledge and employability courses)
- Certificate of Achievement (for students enrolled on a specific integrated occupational programme)
- Certificate of School Completion (for students with significant cognitive delays) (O'Donnell *et al.*, 2012).

The majority of students receive the (Alberta) High School Diploma which is necessary to graduate from high school and, for the purposes of this report, is considered to be the high stakes academic school leaving qualification. Results in the final diploma examination taken on completion of Grade/Year 12 (age 18) influence progression to the various institutions of post-secondary education which exist in the province (O'Donnell *et al.*, 2012).



The marks for the final diploma examinations taken on completion of Grade/Year 12 are determined 50 per cent by class teacher marks, which are submitted to the Ministry, and 50 per cent by the standardised province-wide tests. Tests are usually administered at the end of the school year in June (with retake sessions in January) (O'Donnell *et al.*, 2012).

What format do they take, how does testing take place and what is tested?

High school graduation requirements aim to ensure that students get the basic education they need to do well in their further studies and careers. To receive an Alberta High School Diploma, students must complete a full range of compulsory, core subjects as well as optional courses that broaden their knowledge and skills (O'Donnell *et al.*, 2012).

Diploma requirements are set by the Minister of Education and a student must earn 100 credits in compulsory and optional courses to graduate from high school in Alberta. Schools are required to provide 25 hours of instruction for each credit. Most courses are worth either three or five credits, and the average course load is 35 credits per year.

Grade 12 diploma examinations certify a student's achievement in a course. Each examination is aligned with the programme of studies in order to provide a common assessment for students across the province. This ensures fairness to students and allows them to compete equally for post-secondary admission and scholarships. Students taking the Grade 12 diploma examinations must take courses in the following subjects and combinations (Alberta Education, 2013).

Compulsory	Plus 10 credits in any combination from	Plus 10 credits in any 30-level course which may include:
English language arts	Career and technology studies (CTS)	Locally developed/ acquired and authorised courses
French language arts	Fine arts	Advanced level in career and technology studies courses
Social studies	Second languages	Work experience courses
Pure mathematics	Physical education	Knowledge and employability courses
Applied mathematics	Knowledge and employability courses	
Biology	Registered apprenticeship	
Chemistry		
Physics		
Science		



	Physical education Career and life management	programme courses Locally developed/ acquired and authorised courses in CTS, fine arts, second languages or knowledge and employability occupational courses	Registered apprenticeship programme courses Green certificate specialisation courses Special projects	
--	--	---	---	--

French translations of the diploma examinations in social studies and the mathematics and science courses are provided.

In senior high school, student achievement in each course is normally reported in percentages. The percentages correspond to letter grades as follows:

- A, 80-100 per cent
- B, 65-79 per cent
- C, 50-64 per cent
- F, 0-49 per cent.

To obtain credit in high school courses, students must earn a final mark of 50 per cent or higher. A student who achieves this mark in a given course is eligible to take the next higher course in that sequence. Except for those Grade 12 courses that require provincial diploma examinations, the school determines final marks. For diploma examination courses, the student's final mark is determined by averaging the mark assigned by the school with the mark attained on the diploma examination. This blend of marks provides the most accurate picture of the student's achievement because the two marks are complementary. At the senior high school level, all marks are reported to Alberta Education and become part of the student's record. Alberta Education also issues official transcripts of student achievement (IBE, 2010a and Alberta Education, 2007).

Marking arrangements

Once students have completed an examination, a thorough marking process begins. All examinations and answer sheets (both machine scorable and written answers) are returned to the Assessment Sector of Alberta Education. The answer sheets must be run through scanners and the data checked and verified to ensure there are no errors. After all sheets are scanned, the results are reviewed to make sure that all questions are technically acceptable (Alberta Education, 2013).

Written responses are marked by Alberta teachers (who must hold a valid Alberta teaching certificate, have taught the course for two or more years, and be teaching the course in the school year in which they are selected as a marker). Alberta Education Assessment Sector staff try to ensure a good balance of markers in terms of gender, location and experience. A marking session can last from five to eight



days and require the services of up to 200 teacher-markers who each keep a scoring rubric at their side (Alberta Education, 2013).

The first stage in marking is for a team of experienced teacher-markers (from various parts of the province) to confirm the standards for the written-response marking. These 'standards confirmers' first read sample written responses from previous administrations to ensure that the standards are maintained from semester to semester. They then read a large sample of the current written responses and identify papers that illustrate the standard of expectations at the various scale points in the scoring rubric (e.g., 'excellent', 'proficient', 'satisfactory'). These papers, along with explanatory rationales created by the standards confirmers, are used to train the markers (Alberta Education, 2013).

Each marking day begins with a 'reliability review' at which all markers read a copy of the same paper and score it independently. Once these scores are recorded by a group leader at each marking table, the scores are discussed and differences reconciled. The resulting score of each paper reviewed in this way is posted on the marking floor so that all markers can see the final result (Alberta Education, 2013).

To further enhance fairness, each student's examination booklet is marked by four markers. The first written response is independently scored twice, as is the second written response. None of the markers are aware of the score assigned by any of the other markers. A computer programme then identifies all papers which have a significant discrepancy in scores. These booklets are then given a fifth reading, by another marker, who is unaware of the original scores. The scores this marker assigns are then aligned with the closest original scores to establish a final mark. Should there remain a significant discrepancy between the fifth marker's score and that of the original markers, the paper is then identified for a sixth read to set a final score (Alberta Education, 2013).

Often, but not always, the school-awarded mark is very similar to the diploma examination mark. Cases where diploma examination marks are significantly below teacher-awarded marks are not surprising, since diploma examinations are designed to have an average of 65 per cent – the midpoint between the acceptable standard (50 per cent) and the standard of excellence (80 per cent). Designing the examination to a 65 per cent average does not limit the number of students attaining either standard; it merely ensures the accuracy of the examination for identifying achievement. Should the number of students attaining the standards increase, the provincial average on the test will increase above 65 per cent. Teachers' classroom assessments of students have not changed significantly over time, with averages that have hovered around 75 per cent provincially and a distribution that has remained constant. Examination marks have also been consistent over time, with a mean of approximately 65 per cent, and a similarly constant distribution (Alberta Education, 2013).



What methods are used to establish standards (for each relevant qualification)?

See below.

What methods are used to maintain standards over time (for each relevant qualification)?

In May 2003, Alberta Education announced an initiative designed to maintain, over time, consistent standards associated with diploma examinations. This initiative was introduced to ensure fairness for students, regardless of when they take a diploma examination, and equal opportunity in relation to scholarships and entrance to post-secondary institutions. The initiative aimed to allow the Ministry to report changes in student achievement on diploma examinations from administration to administration and from year to year in a more meaningful manner than had been possible in the past. The initiative also aimed to make it possible to equate students' marks on diploma examinations to accurately reflect their achievement regardless of which form of examination they had taken (Alberta Education, 2008).

To conduct this initiative, it was necessary that at least 20 per cent of the questions on the Part B component of an examination were re-used in another administration. These 'anchor items' are selected from items in previously secured diploma examinations. By comparing student results on anchor items and unique items (unique to a single examination) on any particular diploma examination, and those from the baseline examination, the Ministry can determine whether or not that particular examination was more or less difficult than the baseline examination. Student scores on the particular examination can then be equated to the baseline examination to remove any influence that differences in the difficulty of the two examinations may have had on student scores (Alberta Education, 2008).

Each diploma examination is designed and developed according to a published blueprint that determines the make-up of an entire examination. Once it is established, the blueprint typically remains unchanged through the life of a particular programme of studies, so each examination administered is designed consistently through time. The anchor set of items is selected to be representative of the entire examination, and these items are embedded throughout each examination (Alberta Education, 2008).

When two groups of students take the same set of anchor items contained in two different forms of an examination, the following process occurs. The averages that the two groups attain on the common anchor set are compared to provide information about the nature of the two populations of students. For example, if the averages on the anchor sets for those students taking either form of an examination are almost identical and show no practical significant difference, this tells us that the characteristics of the populations taking the two forms are essentially the same. As a result, any differences in student performance on the unique items of the two forms of an examination would be due to differing item difficulties between the examinations, not differences in the populations taking the examinations. The relative



difficulties of the forms of the examination are then determined (Alberta Education, 2008).

Eight equating methods are used, including equipercentile equating with various smoothing values. Based on the convergence and trends found among these methods, an equating table is developed, which adjusts the students' total scores to the same metric, or standard, found in the baseline examination (Alberta Education, 2008).

This process is not the same as the common notion of scaling examination scores, whereby a uniform amount is added to, or removed from, each examination score. Typically, the numerical adjustments that result from the equating process do not produce a uniform adjustment across all total test scores. It should be noted that test scores of 0 and 100 are never adjusted, because a student cannot have a negative score, nor can a score greater than 100 exist. A student achieving 100 per cent has not necessarily demonstrated an upper limit to his or her achievement, and to move the mark downward would not be fair (Alberta Education, 2008).

Adjustments may be upward or downward dependent upon the differences in difficulty between the baseline examination and the subsequent examination. The degree of adjustment also varies (Alberta Education, 2008).

It is critical to remember that the equating process is designed to remove the possibility of varying examination difficulty as a factor in large-scale assessments. As a result, fairness to all students over time is ensured (Alberta Education, 2008).

The initiative to maintain consistent standards on diploma examinations over time requires that a number of machine-scored questions are unique to each administration and that a number of machine-scored questions (the anchor items) are common across administrations, to allow for comparison of the difficulty of the examinations (Alberta Education, 2008).

Consequently, a score on an examination which is being retaken by an individual candidate may be pro-rated if that candidate has been administered items which they took when they first sat the examination. Students who retake a diploma examination after the baseline examination are scored in the same way as those taking the examination for the first time on the unique items of the examination they are retaking. However, they will have their total score on the Part B component adjusted, based on the ratio of the scores of those taking the total Part B examination for the first time to their scores on the unique items (Alberta Education, 2008).

The process of pro-rating takes into account the fact that the anchor items may differ in average difficulty from the unique items. By pro-rating the scores of those retaking the examinations in this manner, the Ministry can ensure fairness to students who are retaking diploma examinations that contain anchor items, while retaining fairness to those taking the examinations for the first time (Alberta Education, 2008).

The implementation of the initiative to maintain consistent standards on diploma examinations over time has meant that diploma examinations are not released in their entirety immediately after administration. However, the Ministry releases significant numbers of diploma examination items on an annual basis. Booklets of



released items from each of the diploma examinations are provided to schools in print form annually (Alberta Education, 2008).

Where equating is used with a diploma examination, the 'Results Statement' sent to students reports the equated diploma examination mark for each examination taken, as well as the school-awarded mark and the final course mark. In addition, the Results Statement provides the raw score on the written response component of each diploma examination taken (Alberta Education, 2008).

Once the examinations from all administrations in a given school year are equated, results on the multi-year reports will be directly comparable to results in subsequent years (Alberta Education, 2008).

Do monitoring assessments exist?

Alberta Education prepares and administers field tests prior to the January and June diploma examination administration sessions. Field testing is essentially a process of 'testing a test'; test questions are administered to determine their difficulty level and their appropriateness for use on a diploma examination. Teachers are also able to provide comments about potential test questions, thereby contributing to the development of fair, valid and reliable provincial examinations. Teachers are asked to consider the following points when commenting on specific questions and the overall test:

- appropriateness of the questions
- adequacy of writing time limits
- test length
- text readability
- artwork/graphics clarity and appropriateness
- question difficulty.

Field tests may be composed of multiple-choice and/or numerical-response questions, depending on the course and type of field test being administered. Generally, field tests, like diploma examinations, are designed to assess outcomes from entire programmes of study. However, some field tests that Alberta Education administers are 'unit' tests that assess learning outcomes from portions or sections of those programmes of study (Alberta Education, 2013).

What methods are used to review standards, e.g. are there post-examination reviews, and how are the results used?



Finland

Period of compulsory education

Compulsory education starts in the year when a child turns seven and lasts for nine years until age 16. It is provided in a single structure system called basic education. Post-compulsory, upper secondary education is provided in general and vocational upper secondary schools. The general age range for post-compulsory upper secondary studies is 16 to 19 years.

What high stakes school leaving academic qualifications exist and when are they taken?

Lower secondary education

In basic education there are no national examinations; all student assessment is in-school and completed by teachers – either the relevant class or subject teacher. Conduct and schoolwork are assessed by the class teacher or, where a student has several teachers, jointly by these teachers.

Students who have completed the full nine years of compulsory education with a grade of 'adequate' receive the basic education certificate. The final assessment defines how well the student has achieved the objectives of the basic education syllabus (curriculum) in the different subjects. This assessment must be nationally comparable and treat all students equally. Criteria for assessment exist for all core subjects and are set out in the National Core Curriculum (FNBE, 2009).

The scale of numerical grades used in all lower secondary assessment is 4–10, where 5 is adequate, 6 moderate, 7 satisfactory, 8 good, 9 very good and 10 shows excellent knowledge and skills. Grade 4 is for failed performance (FNBE, 2009).

The selection of students for general upper secondary school is based on their grade point average for the academic subjects in the basic education certificate. To be eligible, students should have a minimum grade point average of around 7. Entrance and aptitude tests may also be used, and students may be awarded points for hobbies and other relevant extra-curricular activities. Students must apply for a place at a school providing upper secondary education through a specific application system. All those who have completed the basic education syllabus/curriculum are eligible for general upper secondary education (EACEA, 2013a).

Upper secondary education

At the end of general upper secondary education (at age 19), students usually take the matriculation examination (EACEA, 2013a), the high stakes academic qualification in Finland.

According to the Ministry of Education and Culture's Education and Research Development plan 2007–2012, the objective is for the matriculation examination to measure whether students have acquired the information and skills determined in the upper secondary curriculum and achieved sufficient maturity and eligibility for further



studies in accordance with the objectives of the curriculum (Ministry of Education, Finland, 2007).

The matriculation certification is awarded to a student who has completed the entire general upper secondary school syllabus/curriculum and gives eligibility for entry to higher education. General upper secondary education normally takes three years to complete but, rather than being tied to years/grades (or forms), it is organised into a minimum of 75 courses, each consisting of 38 lessons lasting an average of 45 minutes (IBE, 2010b).

The matriculation examination takes place twice a year, in spring and in autumn, and in all Finnish upper secondary schools at the same time. A candidate has no more than three consecutive examination periods to complete it. The examination can also be completed in a single examination period (EACEA, 2013a).

In addition, students who complete upper secondary education are also issued with the general upper secondary school leaving certificate. This certificate provides details of all the tests passed and grades achieved during upper secondary education (FNBE, 2011).

What format do they take, how does testing take place and what is tested?

The matriculation examination is drawn up nationally and there is a centralised body (The Matriculation Examination Board) to check individual tests against uniform criteria. It is regulated by Section 18 (766/2004) of the Upper Secondary School Act, the Act on the Organisation of the Matriculation Examination (672/2005) and the Government Decree on the Matriculation Examination (915/2005) (FNBE, 2011).

The assessment system for the matriculation examination is different to that used in the rest of the Finnish education system. The Matriculation Examination Board is responsible for administering the examination, its arrangements and execution. The Board issues guidelines on the content, arrangements for and assessment in the matriculation tests. The Ministry of Education nominates the chair of the Board and its members (about 40 in number) at the suggestion of universities, institutions of higher learning and the National Board of Education. The members represent the various subjects covered by the matriculation examination. In addition, around 330 associate members assist the members in the work of preparing and assessing the examinations. The technical arrangements for the examinations are the responsibility of the Board's secretariat, which has 22 civil service employees (FMB, 2007).

The matriculation examination consists of at least four tests/individual examinations; one of which, the examination in the candidate's mother tongue, is compulsory for all candidates. The candidate then chooses three other compulsory tests from among the following:

- an examination in the second national language (Swedish or Finnish – see below)
- a foreign language examination



- a mathematics examination
- and one examination from the general studies 'battery' of examinations (sciences and humanities) (EACEA, 2013a).

The candidate may include, in addition, as part of his or her examination, one or more optional examinations (EACEA, 2013a).

Examinations are offered at two different levels of difficulty in mathematics, the second national language and foreign languages. The levels available in mathematics and foreign languages are the advanced course and the basic course and, in the second national language, the levels available are the advanced course and the intermediate course. The candidate may choose which level of each of the above-mentioned subjects he or she takes, regardless of his or her studies at upper secondary school. The candidate must, however, take at least one examination based on the advanced course in at least one compulsory examination subject. Candidates can take only one examination in the same subject in the matriculation examination (FMB, 2007).

The mother tongue examination is provided in Finnish, Swedish and *Saami*. The Finnish and Swedish tests have two parts: a textual skills section and an essay. In the textual skills test the candidate's analytical skills and linguistic expression are assessed. The essay focuses on the candidate's general level of education, development of thinking, linguistic expression and coherence. The weighted sum of the marks received determines the candidate's grade in the mother tongue examination. In the *Saami* language only an essay test takes place. A candidate whose mother tongue is not Finnish, Swedish or *Saami* can replace the mother tongue test with the Finnish or Swedish second language test (FMB, 2007).

The second national language examinations and the foreign language examinations include sections on listening and reading comprehension and sections demonstrating the candidate's skill in written production in the language in question. The candidate may also take the mother tongue test in Finnish or Swedish in place of the second national language test in that language (FMB, 2007).

In the mathematics examination, the candidate must complete ten questions. He/she is allowed to use a calculator and books of tables that have been approved by the Matriculation Examination Board as aids (FMB, 2007).

General studies options include examinations in Evangelical Lutheran religion, Orthodox religion, ethics, philosophy, psychology, history, social studies, physics, chemistry, biology, geography and health education. The examinations incorporate questions which cross the boundaries of these disciplines. Depending on the examination in question, the candidate answers six or eight test items (FMB, 2007).

The grading system for the matriculation examination is as follows:

L = *laudatur* (Latin for 'praised', 7)

E = *eximia cum laude approbatur* ('passed with exceptional praise, 6)

M = *magna cum laude approbatur* ('passed with much praise', 5)



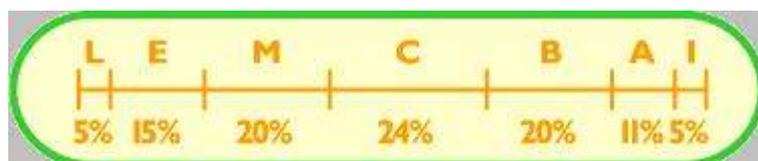
C = *cum laude approbatur* ('passed with praise', 4)

B = *lubenter approbatur* ('satisfactorily passed', 3)

A = *approbatur* ('passed', 2)

I = *improbatur* ('failed', 0) (EACEA, 2013a).

In general, examination grades are allocated as follows:



The relative shares of grades differ somewhat across the various examinations and examination periods.

What methods are used to establish standards (for each relevant qualification)?

Marking

The Matriculation Examination Board organises the marking for all tests. The tests are initially marked in the candidates' schools and then are sent to the Board for centralised marking. A paper by Kaftandjieva and Takala (2002) briefly touches on the marking process for English:

The students' answers are marked first by their own teachers using the official general guidelines and the examination-specific guidelines provided by the English Teachers' Association after consultation with the Board members. Teachers enter their mark on the test papers and on a separate list. All this material is sent to the Board which uses the services of 30–35 experienced raters to mark the English papers centrally...Multiple choice items are handled by optical reading and the score is often weighted...

In the matriculation examination, there are criteria for some subjects. In foreign languages and the mother tongue, the criteria are created according to test type. . In other subjects, the Matriculation Examination Board is developing criteria (Finnish Eurydice Unit, 2013).

What methods are used to maintain standards over time (for each relevant qualification)?

Do monitoring assessments exist?



What methods are used to review standards, e.g. are there post-examination reviews, and how are the results used?

After each examination period, some statistical analysis of the results by test and censors¹⁰ is carried out. Results of this analysis feed into the development of the examinations for the next session. An example of change introduced post review is that, from 2014, relative shares of grades in all subjects will not be used (Finnish Eurydice Unit, 2013).

¹⁰ Censors are the associate members of the Matriculation Examination Board who mark the exams.



France

Period of compulsory education

Compulsory education in France commences at age six and ends at 16. This requirement includes both the *école élémentaire* (elementary school) (five years, ages six-11) and the *collège* (lower secondary school). On average, students leaving the *collège* (which lasts four years, unless they have repeated a year) are 15 years of age. Students must therefore still attend school full-time for at least one more year to satisfy the compulsory schooling requirement, either in a general and technological *lycée* (*lycée d'enseignement général et technologique*, LEGT) or in a vocational *lycée* (*lycée professionnel*, LP). The upper secondary phase, therefore, caters for the final year of compulsory education (students aged 15+) and for the post-compulsory phase, age 16 - 18+ (O'Donnell *et al.*, 2012).

What high stakes school leaving academic qualifications exist and when are they taken?

The *recteur (de l'académie)* (the local representative of the national Ministry of Education) has ultimate responsibility for the setting, organisation and implementation of the national examinations (O'Donnell *et al.*, 2012).

Lower secondary education

At the age of 15, all students take a lower secondary leaving examination. The results of this *diplôme national du brevet* are based on a student's results in this examination, as well as marks for work completed during the two previous years. Passing the *brevet* confirms satisfactory completion of lower secondary education; it does not determine whether students progress to the next stage of schooling at a *lycée* (upper secondary school). Lower secondary examinations are taken in a single session organised each year, on a date set by the Minister for National Education, usually in mid June. Results are sent to parents at the end of the school year (beginning of July) (O'Donnell *et al.*, 2012).

Upper secondary education

The major high stakes academic school leaving qualification is the general *Baccalauréat*. Studies completed in the final two years of general education (in the *première* and *terminale* classes, ages 16-18) lead to a general *Baccalauréat* examination. This is organised to reflect the types of courses offered in the *première* and *terminale* and includes both compulsory and optional examinations. Successful completion entitles the holder to enter higher education (O'Donnell *et al.*, 2012).

There are three types of general (academic) *Baccalauréat*:

- the *BAC L* - literary - main subjects are French, philosophy and modern languages
- *BAC ES* - economic and social sciences
- *BAC S* - scientific - main subjects are mathematics, physics and natural sciences



(O'Donnell *et al.*, 2012).

The *Bac* examinations take place in May and June, with initial results being released in early July (EACEA, 2013b).

Students who complete these studies but do not reach the required standard for the award of a *Bac* may receive the *certificat de fin d'études secondaire*. This certificate is awarded by the *recteur* to those students who do not pass the *Baccalauréat* examination but whose average marks are equivalent to at least 8/20. The certificate states that the student has completed secondary education in its entirety, but it does not entitle the student to enter higher education (O'Donnell *et al.*, 2012).

The other qualifications available are vocational in nature. They include the technological *Baccalauréat*, vocational *baccalauréat/baccalauréat professionnel*, vocational aptitude certificate (*certificat d'aptitude professionnelle - CAP*) and the vocational studies certificate (*brevet d'études professionnelles, BEP*). This desk study does not include further information on these vocational qualifications as its focus is on high stakes academic qualifications (O'Donnell *et al.*, 2012).

What format do they take, how does testing take place and what is tested?

The *Baccalauréat* examinations relate to the official curricula of the final years of classes in *lycées* and are set nationally by the Ministry of Education¹¹ (O'Donnell *et al.*, 2012).

Depending on the specific type of *Baccalauréat* chosen by the candidate, the examinations include compulsory written papers and oral tests and, in some subjects, practical examinations. Full details are available for each type of general academic *Bac* from the Eduscol website:

BAC ES - <http://eduscol.education.fr/cid58532/serie-es.html>

BAC L - <http://eduscol.education.fr/cid58534/serie-l.html>

BAC S - <http://eduscol.education.fr/cid58536/serie-s.html>

The great majority of the examinations are taken at the end of the *terminale* (final year of upper secondary education). However, certain examinations are taken in the penultimate year of upper secondary (the *première*); this is the case with French (EACEA, 2013b).

Marks for the *Baccalauréat* are usually numerical, out of 20. Candidates who receive an average mark equal to or over 10 out of 20 in the final examination session and continuous assessment are awarded a *Baccalauréat*. Those who receive overall marks of less than 8 out of 20 fail. Students who receive average marks of between 8 and 10 for the final examination session and continuous assessment are invited to attend an oral examination at which examiners (the jury) make the decision as to whether students should pass or fail (EACEA, 2013b).

¹¹ Full details – curricula, time allocations etc. are available (in French) from the Eduscol website <http://eduscol.education.fr/cid46205/presentation-du-baccalaureat-general.html>



What methods are used to establish standards (for each relevant qualification)?

The national Ministry of Education is responsible for the education system as a whole. As such, it is responsible for the development, implementation and analysis of monitoring, assessment and evaluation procedures with regard to students and student learning, teaching establishments and education policy and innovation. Comparisons between standards in the different types of *Baccalauréat* are difficult. The examination is marked according to a complicated system, which gives different weightings to results, according to the subject and the type of *Baccalauréat* (general, technological, or vocational, and the specific examination course selected within these types) (O'Donnell *et al.*, 2012).

The *Bac* examinations take months of preparation. A variety of stakeholders are involved in their preparation (Ministry of Education, France, 2012).

The Ministry of Education develops the regulations which define the content and methods of testing for each pathway. It sets the dates for the development process and for the examinations. It shares out the work of developing the examinations between the different *académies* (the local representative areas of the national Ministry of Education) and decides on the amount of resource available for this work (O'Donnell *et al.*, 2012).

The General Inspectorate of Education is statutorily responsible to the Minister for the quality of the *Bac* examinations. Senior inspectors play an important role in the development of examinations, in particular in relation to leadership and guidance (EACEA, 2013b).

The *recteur* of an *académie* is responsible for the organisation of the examinations and the establishment of commissions to develop the papers used for examinations. The development of examinations in the various subject areas begins in the May of the year before the papers are to be sat, so in May 2012, for the examination session in June 2013. The development process is as follows:

- May - Staff at the Ministry of Education, break each subject area down into topics to be developed for next year's examination session. The development of these topic areas is then divided between the *académies*.
- May - in each *académie*, the examination divisions set up 'topic choice commissions' which are presided over jointly by a school inspector and an university faculty member. The members of these commissions are *lycée* (upper secondary) teachers, designated by the *recteur*, and as suggested by the authorised regional pedagogical inspector.
- September to December - each commission develops examination tasks/items for several topics. Each topic item is then submitted separately to two teachers (who have not participated in the work of the commission), who are tasked with completing the examination task developed for a specific topic within a period of time which is less than that to be given to the students who will take the actual examination. These teachers, called 'test teachers', give their opinion on the



feasibility and the interest of the item and detect any errors. The items are modified, if needed, in order to take the test teachers' comments into account.

- December to January - for each topic area, the two co-presidents of the commission prepare and submit a dossier to the *recteur* of the *académie* comprising one item written in its final form, reports from the test teachers and a report from the co-presidents, certifying compliance with current regulations and making suggestions for use of the item. The *recteur* then makes the final choice on the subjects and their assignment.
- March - the document containing the final version of the items is signed by the inspector general and the academic presidents of each commission (Ministry of Education, France, 2012).

There are also juries who play an important role in the grading of students, especially those who have failed the examinations but who may take an oral examination to improve their final result. See above – this relates to those students scoring between 8 and 10 out of 20 in the *baccalauréat* examination (EACEA, 2013b).

What methods are used to maintain standards over time (for each relevant qualification)?

The process described above is intended to ensure that standards are maintained over time.

Do monitoring assessments exist?

What methods are used to review standards, e.g. are there post-examination reviews, and how are the results used?

When and how are the results reported and to whom?



Germany

Period of compulsory education

In the Federal Republic of Germany responsibility for the education system is divided between the Federation and the *Länder* (the administrative states into which Germany is divided). The scope of the Federal Government's responsibilities in the field of education is defined in the Basic Law/ Constitution (*Grundgesetz*). Unless the Basic Law awards legislative powers to the Federation, the *Länder* have the right to legislate. Within the education system, this applies to the school sector, the higher education sector, adult education and continuing education. Administration of the education system in these areas is almost exclusively a matter for the *Länder* (O'Donnell *et al.*, 2012).

Compulsory education lasts from ages six to 18/19. Full-time attendance is compulsory from age six to 15 (nine years) or to age 16 (ten years) (depending on the *Land*). Attendance, at least on a part-time basis, then remains compulsory for a further three years (ages 15/16 to 18/19.) Should students select part-time education from age 15/16 onwards, they usually start out in an occupation and also attend part-time classes receiving some form of vocational training. This is commonly known as the 'dual system' (O'Donnell *et al.*, 2012).

What high stakes school leaving academic qualifications exist and when are they taken?

Lower secondary education

On completion of the courses of education at lower secondary level, students receive one of a range of leaving certificates, provided that they have successfully completed Grade 9 or 10, aged 15/16. In an increasing number of *Länder*, students are required to sit central examinations at *Land* level in order to obtain this leaving certificate. As a rule, students in the *Gymnasium*, which covers both lower and upper secondary level general/academic education, are not issued leaving certificates at the end of the lower secondary level, but receive a qualification to attend the *gymnasiale Oberstufe*, the upper level of the *Gymnasium*. Students who have not achieved the goal of the course of education they were pursuing receive a school leaving report (*Abgangszeugnis*) instead. School leaving certificates and reports are issued by each school and signed by the head teacher and class teacher. The stages of education successfully completed and the qualifications acquired are noted on the leaving certificates (O'Donnell *et al.*, 2012).

Upper secondary education

The *gymnasiale Oberstufe* (upper level of the *Gymnasium* – a secondary school which gives students an in-depth general/academic education aimed at the general higher education entrance qualification) concludes with the *Abitur* examination. Successful completion of the *Abitur* – the high stakes academic qualification in Germany - results in the award of the *Allgemeine Hochschulreife* which is required to



enter higher education (O'Donnell *et al.*, 2012). The *Abitur* examinations are taken on completion of Year 13 in most *Länder*, students aged 18-19. The examinations are designed to test the ability to think analytically rather than to test factual knowledge. Of particular importance are in depth knowledge, skills and competences in German, a foreign language and mathematics. (EACEA, 2013c).

The exact timing of the *Abitur* examinations depends on the individual *Land* but they are usually taken in spring (late March to early June) and results are released in late June or early July (KMK, 2013).

Within the framework of the dual system (work experience combined with in-school education), students in the second year of upper secondary education (usually aged around 17/18) sit an interim examination prior to the end of this year. In technical vocational subjects, such examinations may be practical work tests. Written tasks must be completed to evidence the level of knowledge acquired. Interim vocational examinations are based on the content and capabilities expressed in the appropriate curriculum guidelines. The student receives a certificate confirming his/her participation in the interim examinations and recording his/her standard of education. Various examinations are then taken at the end of the selected course of upper secondary education e.g. vocational examinations or the *Abitur* (O'Donnell *et al.*, 2012).

This study focuses on the *Abitur* as the key high stakes academic qualification which entitles the holder to enter higher education.

What format do they take, how does testing take place and what is tested?

Candidates for the *Abitur* are usually examined in at least four subjects, sometimes five. They take written examinations in their two main subjects; another in which they take written and, in some cases, oral examinations; and a fourth subject which is usually examined orally. Each of the following three subject areas must be represented:

- languages, literature and the arts
- social sciences
- mathematics, natural sciences and technology.

All students must be examined in German or a foreign language (within the subject area languages, literature and the arts) (O'Donnell *et al.*, 2012).

Depending on *Land* legislation, a fifth subject can be examined in either oral or written form, or a particular achievement (*besondere Lernleistung*) which has been performed over at least two half-year terms (e.g. a year paper or the results of a multi-disciplinary project) may be incorporated in the *Abitur* examination. The particular achievement is documented in written form and is complemented by an oral examination (EACEA, 2013c).

To achieve the *Abitur*, students' marks for class work and tests taken in the last two years of school are combined with their results in the *Abitur* examination session to



determine their final result. Schools and subject teachers are responsible for continuous assessment and school examinations (O'Donnell *et al.*, 2012).

What methods are used to establish standards (for each relevant qualification)?

Traditionally, the *Abitur* - the upper secondary leaving certificate/higher education entry qualification - has been set by individual schools/teachers and also marked by students' own teachers, with some papers being verified by the education authority of the *Land* concerned. Although individual schools have traditionally set their own *Abitur* examination, in some *Länder* including Bavaria, Baden-Wuerttemberg, Mecklenburg-Pomerania, Saarland, Saxony, Saxony-Anhalt and Thuringia a centrally-set *Abitur* is used. That is, in these seven *Länder*, schools use an *Abitur* examination drawn up by a commission of teachers and *Länder* Education Ministry officials and not at the individual school level (O'Donnell *et al.*, 2012).

In all *Länder*, it is the *Schulaufsichtsbehörden* (school supervisory authority) that either sets the items for the written examination centrally (that is to say, determines the assignments/assessment tasks), or approves the assessment tasks in cases where these are set by individual schools. As a rule, the school supervisory authority takes advice from teachers and the vast majority of questions are set by the teachers of the respective schools and checked by the school supervisory authority (O'Donnell *et al.*, 2012).

Although usually set and marked at the school level, in accordance with a national agreement of the Standing Conference of the Ministries of Education and Cultural Affairs of all the *Länder*, there are some national guidelines which aim to ensure uniformity in the level of ability required to pass the *Abitur*. The Standing Conference has, for example, adopted standard examination requirements for 33 subjects in the *Abitur*. These detail such items as the tasks to be set in the written examination, the expected written and oral examination performance, or the weighting to be allocated to specific tasks. The *Länder* have, in turn, implemented the decisions adopted by the Standing Conference concerning standard examination requirements for the *Abitur* in the form of *Land* regulations for the written and oral *Abitur* examinations in individual subjects (O'Donnell *et al.*, 2012). These are that there should be:

- an introduction to each of the subjects which contains specific objectives and describes how the subject contributes to the acquisition of basic skills
- a description of the areas of learning and testing at basic and advanced levels to ensure that the subject addresses a range of subject-specific and basic skills
- a description of the performance standards for the subject to enable well-balanced coverage within test questions
- a description of the types of tasks for both oral and written examinations, including examples
- assessment criteria including evidence of what constitutes the mark *gut* (good) (2.0) and *ausreichend* (adequate) (4.0) (O'Donnell *et al.*, 2012).



The Standing Conference's agreement also covers the national six-point marking system which is as follows (O'Donnell *et al.*, 2012):

1	<i>Sehr gut</i> (very good)	Given for a performance which is well above the required standard.
2	<i>Gut</i> (good)	Fully meets the required standard.
3	<i>Befriedigend</i> (satisfactory)	Generally meets the required standard.
4	<i>Ausreichend</i> (adequate)	Shows deficiencies but, on the whole, still meets the required standard.
5	<i>Mangelhaft</i> (poor)	Given when a performance does not meet the required standard, but suggests that the basic knowledge is there and that the deficiencies could be made up in a reasonable period of time.
6	<i>Ungenügend</i> (very poor/ unsatisfactory)	Performance does not meet the required standard. Basic skills are so incomplete that the deficiencies could not be made up in a reasonable period of time.



Standardisation of marking

The results of the various written assessment assignments undertaken provide teachers with feedback on the success of their teaching and a basis for proceeding further. *Länder* with central examinations are able to monitor this feedback system closely, since each teacher has an anonymous co-marker (O'Donnell *et al.*, 2012).

Issues

The *Abitur* qualification is recognised across Germany; as a result, students leaving secondary school with the qualification can study in any *Land*. Statistics show, however, that there are large, sometimes even dramatic differences between the *Länder*. While in Saarland 50.7 per cent of an age group attains the general qualification for university entrance, this figure is only 25 per cent in Bavaria. The average final grades are also quite disparate – in 2005, the average mark for the *Abitur* examination in Thuringia was 2.30; in Lower-Saxony it was 2.72. These numbers could suggest variability of degree of difficulty in the examinations in the different *Länder* or different levels of performance on part of the school systems or students. Either way, they indicate a deficit with regard to equal opportunity in comparison to other countries, especially since students compete for admission with applicants on a national and even international scale (Goethe Institute, 2011).

Reforms

The Standing Conference of the Ministers of Education and Cultural Affairs of the *Länder* in the Federal Republic of Germany is seeking to improve the comparability of the general higher education entrance qualification (*Allgemeine Hochschulreife*) in Germany by establishing educational standards in key subjects; offering a national pool for *Abitur* questions; and establishing uniform assessment criteria (O'Donnell *et al.*, 2012).

In an initial step, in autumn 2012, standards were adopted for German, mathematics, English and French. Work on developing educational standards for the *Abitur* in the natural-science subjects - biology, physics and chemistry will commence in 2013.

The standards are available on the KMK website

<http://www.kmk.org/dokumentation/veroeffentlichungen-beschluesse/bildung-schule/qualitaetssicherung-in-schulen.html#c7035>.

In further steps, the Institute for Educational Progress (*Institut zur Qualitätsentwicklung im Bildungswesen* - <http://www.iqb.hu-berlin.de/> (IQB)) has started to develop model questions for the *Abitur* examinations, including level of expectation, as well as information on how to assess the answers. The *Länder* are supporting this process by submitting further written *Abitur* examination questions, as well as relevant assessment criteria, to the committee responsible for developing examination questions. The *Abitur* questions that the IQB deems suitable will then be included in a pool of questions. Further information about the 'taskpool' is available [online](#). (IQB, 2012).

All *Länder* have committed to implementing the educational standards for upper secondary education from 2013/2014. From the 2016/17 school year, the final



examinations in all *Länder* should be based on the new educational standards (KMK, 2012).

A Powerpoint presentation (in German) is available which discusses the development of these standards http://www.iqb.hu-berlin.de/bista/abi/PK_2012.09.19_FI.pptx

What methods are used to maintain standards over time (for each relevant qualification)?

Responsibility for arrangements to monitor and assess performance lies with the highest authority with responsibility for educational policy in each *Land* (Ministries, Senates). In *Länder* where there is a centralised final *Abitur* examination, the implementation and standard of this examination is again determined by the highest authorities in the *Land* (O'Donnell *et al.*, 2012).

Do monitoring assessments exist?

The 'Uniform Exam Conditions' (*Einheitliche Prüfungsanforderungen*, EPA) state that examination standards are open to development and should be checked/monitored and further developed at the appropriate time (KMK, 2008).

The model questions for the *Abitur* examinations that are planned to be made available to *Länder* from 2016/17 are being reviewed by two different working groups:

- The *Abitur* Commission (*AGs Abiturkommission*) composed of *Länder* and IQB representatives, which is responsible for policy decisions that set the framework for the development and use of the 'taskpools' and the general criteria for task selection.
- The Task Evaluation group (*AGs Aufgabenbewertung*), comprising technical specialists from the *Länder* and IQB and also education researchers and professionals, which specifies the criteria for the selection of tasks in individual subjects. The criteria relate specifically to the standard terms for the tasks and the adequacy of the required levels (IQB, 2012).



Hong Kong

Period of compulsory education

A new structure for compulsory education was introduced in September 2009. Education in Hong Kong now comprises six years of primary education, three of lower secondary, and three of upper secondary education, with the primary and lower secondary phases being compulsory, that is, age six to age 15. Previously, the structure of the education system reflected the English model, with six years of primary (divided into two, three-year cycles in Hong Kong), three years of lower secondary, two years of senior secondary and two years of post-secondary education (IBE, 2010c).

What high stakes school leaving academic qualifications exist and when are they taken?

Students in post-compulsory upper secondary education (age 15 to 17) prepare for the Hong Kong Diploma of Secondary Education (HKDSE), the examinations for which are held at the end of Form 6 (age 17). The diploma has four core subjects – English, mathematics, Chinese and liberal studies. Students then choose two or three elective subjects from a choice of 20. There are also some applied learning subjects, modelled on the idea of the BTEC qualification in England, and six other modern foreign languages which can also form part of the students' choices (ITS, 2013).

The examinations are taken in late March, April and early May. Results are released in July (HKEAA, 2013a).

The HKDSE examination is administered by the [Hong Kong Examinations and Assessment Authority](#) (HKEAA).

The first candidates taking the HKDSE completed it in 2012. It replaced both the Hong Kong Certificate of Education Examination (HKCEE) and the Hong Kong Advanced Level Examination (HKALE).

Results in the HKDSE are expressed in terms of five levels of performance, of which 5 is the highest and 1 the lowest. The Level 5 candidates with the best performance are awarded a 5**, and the next top group awarded a 5*. A performance below Level 1 is labelled as "unclassified" (HKEAA, 2013b).

The HKDSE allows holders to enter higher education, enrol on post-secondary courses (including programmes leading to Associate Degrees, Higher Diplomas and Certificates), and/or enter vocational education and training. It is the new benchmark examination of the Joint University Programmes Admissions System (JUPAS, the undergraduate admission allocation system in Hong Kong) (HKEAA, 2013b).



What format do they take, how does testing take place and what is tested?

The HKDSE combines the results of public examinations and school-based assessments.

Assessments in the HKDSE (both public examinations and school-based assessments) are aligned with the aims, objectives and intended learning outcomes of the senior secondary curriculum, details of which are available [online](#).

Depending on the subject, various items, including multiple-choice questions, short questions and long questions, are used in public examinations to assess students' performance in a broad range of skills and abilities. Multiple-choice questions permit a more comprehensive coverage of the curriculum; short questions can be used to test basic knowledge and concepts. Longer questions aim to test candidates' higher-order skills (HKEAA, 2008).

Most school candidates take four core subjects (Chinese language, English language, mathematics and liberal studies), plus two or three elective subjects from the new secondary structure (category A), applied learning subjects (category B) and other language subjects (category C) (HKEAA, 2008).

There are 20 elective subjects available under category A; they are:

- biology
- business, accounting and financial studies
- chemistry
- Chinese history
- Chinese literature
- design and applied technology
- economics
- ethics and religious studies
- geography
- health management and social care
- history
- information and communication technology
- literature in English
- music
- physical education
- physics
- science
- technology and living



- tourism and hospitality studies
- visual arts (HKEAA, 2008).

The applied learning subjects (category B) include:

- design studies
- media arts
- performing arts
- films, TV and broadcasting studies
- media writing and production
- business studies
- clientele/customer management
- legal studies
- hospitality services
- event management
- personal and community services
- medical science and health care
- sports
- civil and mechanical engineering
- services engineering (HKEAA, 2008).

For the other languages subjects (category C), the Cambridge International Examination AS question papers are used (HKEAA, 2008).

This description focuses on the arrangements for the core subjects and category A courses.

School-based assessment (SBA) refers to assessments administered in schools and marked by the students' own teachers. SBA marks count towards students' results for the HKDSE.

The primary rationale for SBA is to enhance the validity of the assessment, by including the assessment of outcomes that cannot be readily assessed within the context of a one-off public examination. SBA can also reduce dependence on the results of public examinations, which may not always provide the most reliable indication of the actual abilities of candidates. The results of assessments based on student performance over an extended period of time and developed by those who know the students best - their subject teachers – is believed to offer a more reliable assessment of each student. Another reason for including SBA is to promote a positive impact on teaching and learning. It can serve to motivate students by engaging them in meaningful activities; and for teachers, it can reinforce curriculum aims and good teaching practice, and provide structure and significance to an activity that they are in any case involved in on a daily basis, namely assessing their own



students (HKEAA, 2009).

Of the 20 elective category A subjects, only 14 currently have SBAs. This will be extended to all subjects for HKDSE in 2014 (HKEAA, 2009).

The modes of SBA vary across subjects but they all form an integral part of the learning and teaching process. Details are provided – for the core subject areas - in the table below: (HKEAA, 2009)

Chinese language	The SBA includes reading activities, coursework/other language activities and assessment of the elective modules. Students' learning outcomes are assessed by various assignments, written reports and/or oral presentations.
English language	Students read/view a number of texts (including books, films and documentaries) over the course of study and study three elective modules. They take part in group discussions and make individual presentations based on what they have read/viewed, and what they have learned in the elective modules.
Liberal studies	Students complete a piece of independent study/enquiry, for which they formulate a project title, collect relevant data and prepare a project report after analysing the data collected.
Elective subjects	Assessment activities for elective subjects are integral to subject learning i.e. the activities are normal in-class/out-of-class activities that students are expected to be involved in during the learning process, e.g. laboratory work in science subjects, project work in information and communication technology, and portfolios in visual arts.

The HKEAA ensures that the following measures are in place to ensure fairness in SBA marking:

- Detailed guidelines, assessment criteria and exemplars are provided to ensure consistency in teachers' assessments.
- Professional development training is provided to help teachers become familiar with how to conduct the SBA in their subject(s).



- District coordinators are appointed to support schools in the conduct of the SBA for individual subjects.
- SBA marks submitted by different schools are moderated to iron out possible differences between schools in marking standards (HKEAA, 2009).

What methods are used to establish standards (for each relevant qualification)?

A standards-referenced reporting system – for the HKDSE - has replaced the reporting system used for HKCEE and HKALE. Standards-referencing aims to report candidates' results against a set of prescribed levels of achievement based on the typical performance of candidates at those levels. For each of the levels, a set of descriptors has been developed that describes what the typical candidate performing at this level is able to do. Candidates' results indicate the standards achieved in terms of knowledge and skills, regardless of the performance of the other candidates taking the same examination. This reporting system also enables stakeholders to understand explicitly what the candidates know and can do when they have achieved a certain level of performance. The results are expressed in terms of five levels of performance, of which 5 is the highest and 1 the lowest. The Level 5 candidates with the best performance are awarded a 5**, and the next top group awarded a 5*. A performance below Level 1 is labelled as "unclassified" (HKEAA, 2013b).

In terms of these new level descriptors, performance standards at Levels 4 and 5 have been prepared with reference to the standards achieved by candidates awarded grades D or above in the previous Hong Kong Advanced Level Examination (HKALE). This referencing is important to ensure a degree of continuity with past practice, to facilitate tertiary selection and to maintain international recognition. The standards are illustrated with exemplars of students' work at the various levels of attainment. The exemplars, together with the level descriptors, clarify the standards expected at each individual level (HKEAA, 2011).

The standards are applied by an expert panel of judges for each HKDSE subject in such a way that the 'cut-scores' for each level represent the same level of achievement each year. Different types of evidence including data from monitoring tests, the level descriptors, candidates' current and past levels of performance and general statistical data are considered in this process. The recommended cut-scores are approved by the Public Examinations Board (PEB) before being applied to the generation of examination results (HKEAA, 2011).

The above mechanism aims to ensure that no single factor or professional can predominate in the process, and the standards can be maintained and held constant without 'grade inflation' over time (HKEAA, 2011).

The level descriptors for the HKDSE are as follows (HKEAA, 2011):



Candidates at this level typically demonstrate:	
Level 5	<p>Comprehensive knowledge and understanding of the curriculum and the ability to apply the concepts and skills effectively in diverse and complex unfamiliar situations with insight</p> <p>Ability to analyse, synthesise and evaluate information from a wide variety of sources</p> <p>Ability to communicate ideas and express views concisely and logically</p>
Level 4	<p>Good knowledge and understanding of the curriculum and the ability to apply the concepts and skills effectively in unfamiliar situations with insight</p> <p>Ability to analyse, synthesise and interpret information from a variety of sources</p> <p>Ability to communicate ideas and express views logically</p>
Level 3	<p>Adequate knowledge and understanding of the curriculum and the ability to apply the concepts and skills appropriately in different familiar situations</p> <p>Ability to analyse and interpret information from a variety of sources</p> <p>Ability to communicate ideas and express views appropriately</p>
Level 2	<p>Basic knowledge and understanding of the curriculum and the ability to apply the concepts and skills in familiar situations</p> <p>Ability to identify and interpret information from straightforward sources</p> <p>Ability to communicate simple ideas in a balanced way</p>
Level 1	<p>Elementary knowledge and understanding of the curriculum and the ability to apply the concepts and skills in simple familiar situations with support</p> <p>Ability to identify and interpret information from simple sources with guidance</p> <p>Ability to communicate simple ideas briefly</p>
To ensure that there is no mismatch between students' abilities and the anticipated standards, when developing the HKDSE, sample papers were piloted on senior	



secondary students with different abilities from a range of schools. The performance sample collected was analysed and used as the basis for the development of the level descriptors. Where necessary, performances at different grades in past HKALE and HKCEE papers were also referred to (HKEAA, 2011).

Moderation for school-based assessment (SBA)

The moderation process deals with teachers' assessment results for their students (that is, the marks awarded to students by teachers) and takes place after the assessment is completed and marks have been submitted to the HKEAA. The main reason for this moderation is to ensure the fairness of the SBA. Teachers know their students well and are consequently well-placed to judge their performance. In consultation with their colleagues, they can reliably judge the comparative performance of all students within the school in a given subject. However, they are not necessarily aware of the standards of performance across all schools. Despite the training provided to teachers in carrying out SBA, and even given that teachers assess students on the same tasks using the same assessment criteria, teachers in one school may be harsher or more lenient in their judgements than teachers in other schools. To address these potential problems, the HKEAA moderates SBA marks as appropriate, and makes use of a combination of methods to moderate or adjust the assessment scores submitted by schools, with the aim of ensuring the comparability of SBA scores across schools. These methods make combined use of both statistical models and expert judgement moderation (HKEAA, 2008).

Statistical models are used to adjust the average and the spread of SBA scores of students in a given school with reference to the public examination scores of the same group of students. Students' SBA marks may be adjusted but the rank order determined by the school remains unchanged. The HKEAA collects samples of students' work for inspection, and follows up on any outliers identified from the inspection of samples in the moderation process with appropriate action (HKEAA, 2008).

HKEAA-appointed professionals (e.g. District Coordinators, Assessors) conduct inspections of samples of students' work covering the full range of attainment, and make recommendations for mark adjustments. The HKEAA specifies a sufficiently large sample size for inspection, which ensures a good evaluation of a school's judgement at different performance levels. Additional samples may be requested if necessary (HKEAA, 2008).

For more details on SBA moderation, see (HKEAA, 2008.)

What methods are used to maintain standards over time (for each relevant qualification)?

In the development of public examinations by the HKEAA, rigorous quality assurance practices are strictly observed. Key elements of the HKDSE quality assurance practices are:

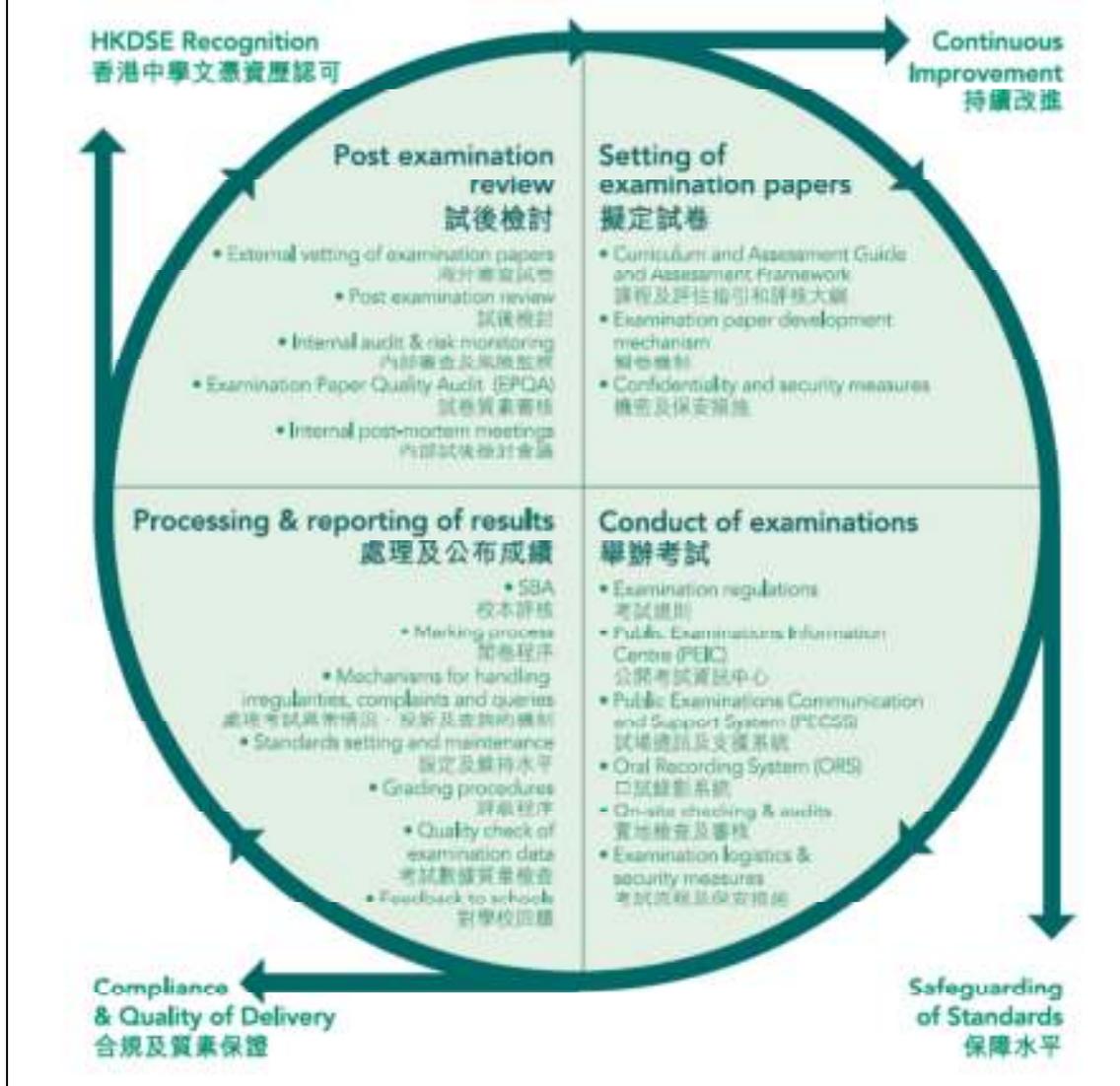
- Standards' setting - Standards are defined by level descriptors, and the standards are illustrated with exemplars of students' work at the various levels of



attainment. These standards are applied by an expert panel of judges for each HKDSE subject and the 'cut-scores' for each level represent the same achievement each year. The recommended cut-scores are approved by the Public Examinations Board before being applied to generate examination results. Further details of the Grading Procedures and Standards-referenced Reporting are available at <http://www.hkeaa.edu.hk/en/hkdse/SRR/>.

- Moderation of school-based assessment (SBA) results - To ensure the reliability and consistency of assessment standards, school-based assessment (SBA) marks are moderated as appropriate. A combination of methods is used, including statistical moderation and expert judgement, to moderate or adjust the assessment scores submitted by schools, with the aim of ensuring the comparability of SBA scores across schools.

The diagram below, from the HKDSE Examination Quality Assurance Framework Booklet ([HKEAA, 2011](#)) provides an overview of the quality assurance framework for the HKDSE examination.



Do monitoring assessments exist?

In order to provide some background information about standards in the HKDSE and previous HKCE/HKAL examinations, secure monitoring tests for the core subjects were administered to samples of HKCEE school candidates in 2010 and HKALE school candidates in 2011. The same monitoring tests were administered to a selected sample of candidates sitting the 2012 HKDSE examination. It was intended that, in the first administration, the statistical data from the monitoring tests would help provide some linkage of the standards between the HKDSE and the previous HKCE and HKAL examinations.

What methods are used to review standards, e.g. are there post-examination reviews, and how are the results used?

When and how are the results reported and to whom?

A variety of post-examination review methods are in place in Hong Kong. These include:

- External vetting of examination papers - samples of candidates' scripts in HKDSE subjects, together with the question papers and marking schemes, are sent for overseas vetting annually. The main purpose of this is to collect feedback and comments on the examination papers and marking standards, in order to identify opportunities for improvement, and also to ensure that HKDSE standards are comparable with those of overseas examinations, so that international recognition can be sustained.
- Each year, after the examination, subject committees review the examination papers and marking schemes in their regular meetings. This monitoring mechanism is important as it can help to ensure continuous improvement in the quality of the examination papers and marking. Different aspects regarding the examination papers and marking schemes are reviewed, including the aims and objectives, coverage, standards, balance of basic knowledge / skills versus higher-order skills, wording, clarity and appropriateness of the marking scheme, choice of questions, comparison of language versions, etc. Where applicable, the Chief Examiner's comments and any complaints from schools and individuals, are also discussed with a view to identifying areas for improvement.
- Internal audit and risk monitoring - to ensure that the practices documented in guidelines and instructions are complied with, internal audits are regularly conducted in the HKEAA, in accordance with the requirements of International Standard ISO 9001 and the risk-based internal auditing approach. Key risks associated with the provision of the HKDSE examination are monitored and reviewed regularly, in accordance with HKEAA's risk management system. Relevant risk mitigation measures (e.g. contingency planning / fallback arrangements) are in place with the aim of minimising the possible impacts resulting from any unexpected events, and in order to ensure that risks are managed at an acceptable level, and continuous improvement of examination



processes can be achieved.

- Examination Paper Quality Audit (EPQA) - subject managers perform post-examination reviews on the paper design and implementation after the conduct of the examination. These reviews can be further formalised and enhanced from a psychometric perspective. Appropriate techniques for analysing examination data allow the review of the psychometric properties of an examination paper, such as its validity, reliability and equitability. It is also possible to gauge how well individual items have worked. Furthermore, the computation steps for scoring and scaling to determine 'cut-scores' are reviewed. If there is a pilot test of the test items, the statistical aspects of the design of the pilot test and use of the pilot results are also examined. The EPQA includes scrutiny of the quality of the examination papers (based on subject expertise), analysis of examination response data and study of statistical and computational procedures.
- Internal 'post-mortem' meetings - post-mortem meetings are conducted annually after the examination cycle, at both divisional level and cross-divisional level in the HKEAA. At the meetings, notable issues arising from different stages of the public examinations are discussed and reviewed, with the aim of identifying areas of improvement for the future (HKEAA, 2011).



Massachusetts - USA

Period of compulsory education

Compulsory education in Massachusetts last for ten years between the ages of six and 16 (ECS, 2010). Each school committee may establish its own minimum permissible age for school attendance, provided that this age is not older than the mandatory minimum age. The final years of compulsory upper secondary schooling necessary for graduation are provided in *high school*, usually including Grades 10-12 (students aged 15-18) or Grades 9-12 (students aged 14-18) (O'Donnell *et al.*, 2012).

What high stakes school leaving academic qualifications exist and when are they taken?

Assessment in the United States is primarily undertaken at state and local levels. Most states set standards for what students should learn and how well they should learn it (O'Donnell *et al.*, 2012).

The key assessment system in Massachusetts is the Massachusetts Comprehensive Assessment System (MCAS) which:

- tests all public school students across the state, including students with disabilities and students with limited English proficiency
- is administered annually
- measures performance based on the Massachusetts curriculum framework learning standards
- reports on the performance of individual students, schools, and districts
- serves as one basis of accountability for students, schools, and districts (for example, since 2003, students in Grade 10 have had to pass the Grade 10 MCAS tests as one condition of eligibility for a high school diploma - normally received on completion of Grade 12, aged 18).

The Massachusetts Comprehensive Assessment System (MCAS) was implemented in response to the Education Reform Act of 1993. It was originally implemented for students in Grades/Years 4, 8, and 10 (aged around ten, 14 and 16 respectively) and has since been expanded to include children in Grades 3, 5, 6 and 7 in addition (so, it now applies in each year group/Grade from age nine to age 14, and at age 16).

MCAS testing usually takes place in late spring (April/May), with scoring taking place in July/August.

To be eligible for a high school diploma (achieved on completion of Grade 12, aged 18), students are required to pass the MCAS Grade 10 tests in English language arts and mathematics and one high school test in one of biology, chemistry, introductory physics, or technology/ engineering. Students must also fulfil any specific local requirements in addition (this could include minimum years of study per subject or number of credits achieved per subject, and can be specific to each district or school



within Massachusetts). Students are provided with multiple opportunities, if necessary, to pass the Grade 10 tests (MDESE, 2013).¹²

In addition, the MCAS programme is used to hold schools and districts accountable, on an annual basis, for the progress they have made towards the objective of the No Child Left Behind Act that all students are proficient in reading and mathematics by 2014 (MDESE, 2013).

The academic standards introduced under the Education Reform Act of 1993 include criteria for three determinations of certificates of student academic performance during education in Massachusetts (to Grade 12, age 18), as follows:

Competency determination

This is an assessment of all Grade 10 students to determine whether they have 'demonstrated mastery of a common core of skills, competencies and knowledge' in the areas of mathematics, science and technology, history and social science and English. Students must pass the MCAS to achieve this; achieving this is also a condition of achieving a high school diploma.

Certificate of Mastery

This designation is accorded to high school students who demonstrate mastery of a 'comprehensive body of skills, competencies and knowledge' comparable to that possessed by high school graduates 'in the most advanced education systems in the world'. The criteria for a Certificate of Mastery may incorporate a number of factors which may include, but not be limited to, any of the following: high school graduation standards, superior performance on Advanced Placement (AP) tests administered by the Educational Testing Service (ETS), and demonstrated excellence in areas not reflected by the state's assessment instruments, such as artistic or literary achievement.

Certificate of Occupational Proficiency

This designation is conferred on students who complete a 'comprehensive education and training programme in a particular trade or professional skill area' provided they demonstrate a mastery of core skills, competencies and knowledge comparable to students entering the trade or profession in the 'most educationally advanced education systems in the world'. To receive this certificate a student must first have acquired a competency determination (Commonwealth of Massachusetts, 2013).

¹² The Massachusetts Department of Elementary and Secondary Education does not rank cities or towns based on [MCAS scores](#). While local media use statewide results to create their own rankings, this practice is not encouraged or endorsed by the Department of Elementary and Secondary Education.



What format do they take, how does testing take place and what is tested?

Massachusetts Comprehensive Assessment System (MCAS) testing involves the following types of assessment:

- Multiple-choice questions: these are used in all content/subject area tests and students select an answer from four options
- Short-answer questions: these are used in mathematics tests only. Students generate a brief response, for example, a short statement or computation leading to a numeric solution
- Open-response questions: these are used in all content area tests. Students create a one- or two-paragraph response in writing, or in the form of a narrative or a chart, table, diagram, illustration, or graph, as appropriate
- Writing prompts.

As a result of the (Massachusetts) Education Reform Act of 1993, curriculum frameworks for pre-kindergarten to Grade 12 (high school) students have been established to cover seven discipline areas:

- mathematics
- science and technology/engineering
- history and social science (includes US and world history, geography, economics, civics and government)
- English language arts
- foreign languages
- the arts (includes dance, music, theatre and the visual arts)
- health (includes health education, physical education and family and consumer science education) (MDESE, 2011a).

Writing prompts are only used in English language arts tests.

What methods are used to establish standards (for each relevant qualification)?

The MCAS programme is managed by Massachusetts Department of Elementary and Secondary Education staff with assistance and support from an assessment contractor, Measured Progress. The contract with Measured Progress involves complete operational implementation of assessments in English language arts, mathematics, and science and technology/engineering. The MCAS tests more than 500,000 students per year in Grades 3 through to 8 (and in Grade 10) in high school.

Massachusetts educators play a key role in the MCAS through service on a variety of committees related to the development of MCAS test items, the development of



MCAS performance-level descriptors, and the setting of performance standards. The programme is supported by a five-member national Technical Advisory Committee as well as measurement specialists from the University of Massachusetts–Amherst (MDESE, 2011b).

Content standards are developed for each of the seven curriculum framework subject areas. Each content standard has four components:

- Core concepts: these are 'organisers' or 'big ideas' which have been designed to help teachers and students conceptualise learning in the discipline
- Strands: these describe the broad areas of knowledge and skills that students encounter pre-kindergarten through to Grade 12 (age 18) as they study the core concepts of a discipline
- Learning standards: these articulate the specific content that students study and the skills through which students practise the strands in developmentally appropriate ways. These are presented in Grade level clusters: pre-kindergarten to Grade 4 (age 10), Grades 5-8 (ages 10-14), Grades 9 and 10 (ages 14-16); and Grades 11 and 12 (ages 16-18)
- Examples of student learning: these classroom ideas help teachers identify how the curriculum can be structured for students to demonstrate increasing mastery of content and skills at the Grade level clusters.

Each curriculum framework is always considered to be work in progress and, consequently, is subject to continual refinement and strengthening. The frameworks are also continually reviewed to ensure that they remain current. Local communities use the frameworks to develop more specific, locally-adapted curricula, and the Department of Education bases the statewide student assessment programme on the frameworks. These standards apply to all grades in Massachusetts from pre-kindergarten through to high school (O'Donnell *et al.*, 2012).

Whilst the (Massachusetts) Education Reform Act of 1993 required the state to establish educational goals, academic standards, curriculum frameworks and a system for evaluating individual schools and school districts and, very specifically required the state Board of Education to develop academic standards in the seven core subject areas, the Act also gave the Board discretion to develop academic standards in other subject areas if it chose to do so.

The Education Reform Act of 1993 specifically directed that any standardised tests used for assessment should be criterion-referenced in order to assess whether students are meeting the academic standards set out in the curriculum frameworks. For students whose performance is difficult to assess, the statute recommends the development of alternative methods such as work samples, projects and portfolios.

Massachusetts Comprehensive Assessment System (MCAS) results are reported for individual students, schools and districts according to four performance levels defined by the Board of Education:

- advanced



- proficient
- needs improvement
- warning/failing.

MCAS is externally marked by professional scorers and Massachusetts teachers who have been specifically trained. Responses to open-response questions are scored, by professional scorers only, using a scoring guide or rubric. MCAS rubrics indicate what knowledge and skills students must demonstrate to earn 1, 2, 3 or 4 score points.

Student compositions are scored by trained teachers and are evaluated on two criteria:

- topic development, based on a 1-6 score point scale
- standard English conventions, based on a 1-4 score point scale.

For all but long compositions, scorers view electronic copies of digitally scanned images of student responses on a computer monitor and assign scores electronically. Use of this computerised scoring system aims to ensure that student responses are randomly assigned to scorers; that second readings are truly 'blind'; and that supervisors have immediate access to information about scorer accuracy. Students' long compositions are each scored twice (by two separate scorers) for both topic development and standard English conventions.

What methods are used to maintain standards over time (for each relevant qualification)?

Technical characteristics of the internal structure of the assessments are measured in terms of classical item statistics (item difficulty, item-test correlation), differential item functioning analyses, dimensionality analyses, reliability, standard errors of measurement, and item response theory (IRT) parameters and procedures. Each test is equated to the previous year's test in that Grade/Year and content area in order to preserve the meaning of scores over time (MDESE, 2011b).

Do monitoring assessments exist?

The Center for Educational Assessment has been working with the Massachusetts Department of Elementary and Secondary Education over the past ten years to prepare reports that address the validity of different aspects of the Massachusetts Comprehensive Assessment System (MCAS). (University of Massachusetts Amherst, 2011)

The first study explored items administered in the 2008, 2009 and 2010 Grade 8 science, technology and engineering (STE) assessments. A similar study was conducted on the 2008, 2009 and 2010 Grade 10 English language arts (ELA) assessments. Both studies concluded by remarking that any advantages in favour of one sub-group over another were small or non-existent, thus furthering the validity evidence (MDESE, 2011b).



What methods are used to review standards, e.g. are there post-examination reviews, and how are the results used?

Technical reports are published each year to evaluate test development, they are available here: <http://www.mcasservicecenter.com/McasDefault.asp?ProgramID=31>

They address:

- test administration
- equating and scaling of tests
- statistical and psychometric summaries
 - item analyses
 - reliability evidence
 - validity evidence.

The purpose of the technical reports is to document the technical quality and characteristics of the MCAS operational tests, to present evidence of the validity and reliability of test score interpretations, and to describe modifications made to the programme during that year (MDESE, 2011b).



The Netherlands

Period of compulsory education

The obligation to attend school is laid down in the Compulsory Education Act. Every child must attend school full time from the first school day of the month following his/her fifth birthday. In reality, however, nearly all children attend school from the age of four. Children must attend school full time for 12 full school years and, in any event, until the end of the school year in which they turn 16 (EACEA, 2013d).

Under the basic qualification requirement that came into effect in September 2007, all young people up to the age of 18 must also attend school until they attain a basic qualification. A basic qualification is a HAVO, VWO or VMBO qualification. Young people under the age of 18 who are no longer in full-time education are required to follow a full-time programme combining work and study, such as block or day release, until they have obtained one of the required certificates (EACEA, 2013d).

In curricular terms, secondary education is divided into three distinct types, named according to the final qualification for which each type prepares its students:

- VMBO: four-year course for students aged 12-16, preparing them for the pre-vocational education certificate (as a basis for further vocational training/study)
- HAVO: five-year course for students aged 12-17, which leads to the general upper secondary education certificate, and prepares students for higher professional education. A HAVO certificate also allows students to go on to pre-university education
- VWO: six-year course, for students aged 12-18, leading to the pre-university certificate, which is designed to prepare students for university. There are two main types of VWO school - the *gymnasium* and the *atheneum*.

The VWO is the high stakes general/academic school leaving academic qualification and, consequently, is the focus for this study (O'Donnell *et al*, 2012).

What high stakes school leaving academic qualifications exist and when are they taken?

After six years of pre-university education, students sit the examinations leading to the award of the VWO which grants access to university, higher professional education (HBO) or employment.

The VWO school leaving examinations are in two parts: a national examination held in the final year and a component organised by the school, known as the school examination. The elements of the school examination must be completed before the national examination begins.

The national examinations take place in May, with opportunities for resits in June and August. (Government of the Netherlands, 2012)

VWO students choose from four fixed subject combinations:



- science and technology
- science and health
- economics and society
- culture and society.

Each group of subjects includes:

- a common component, which accounts for 35 to 43 per cent of the curriculum
- a specialised component (consisting of subjects relating to the chosen subject combination), accounting for 39 to 48 per cent of the curriculum
- an optional component accounting for 18 to 20 per cent of the curriculum (students are free to choose from the subjects offered by the school, including subjects provided through an arrangement with other schools; the number of optional subjects depends on the study 'load' in the specialised component).

A revised curriculum was launched on 1 August 2007 for all students entering the fourth year of VWO. Following this, some examination syllabuses were completely revised (e.g. history, geography and social studies) and others were radically rearranged (science, life and technology and mathematics). The first national examinations for the VWO under the new system were held in 2009/2010 (EACEA, 2013d).

The national examination syllabuses for history, geography and social studies were completely overhauled while few, if any, changes were made to the remaining syllabuses. The old and new systems do, however, differ fundamentally for all subjects as regards the descriptions contained in the national examination syllabuses. There are other differences, too:

- The syllabuses are shorter and less detailed
- The national examination no longer covers the entire syllabus. This applies to nearly all subjects
- Schools can add their own elements to those described in the national examination syllabus
- The new syllabuses no longer prescribe how elements must be tested or how they must be weighted.

What format do they take, how does testing take place and what is tested?

The VWO school leaving examinations are in two parts: a national examination held in the final year and a component organised by the school, known as the school examination. For some subjects there is a school examination only.

The school examination usually takes the form of an examination portfolio comprising



various elements as documented in a form decided upon by the school, for example a list of grades/marks or examples of project work. The requirements to be met by the school examination are set out in the examination syllabus and cover all the elements that make up the examination portfolio for each subject. The separate elements of the school examination are not all scheduled for the final year. Each school can decide when the various parts of the examination are to be held.

Every year, schools are required to submit their own school examination syllabus to the Inspectorate showing what elements of the syllabus are tested when, and how marks are calculated, including the weight to be allocated to the different tests, and opportunities to resit them. To help schools, the SLO (the National Institute for Curriculum Development) publishes school examination guidelines for every subject and level of education. These guidelines are not compulsory, in contrast with national examination syllabuses, but can help schools set examinations (EACEA, 2013d).

SLO guidance for schools is [available](#).

In general, the school examination consists of two or more tests per subject. These may be oral, practical or written. There are also practical assignments for which no marks are given, just an acknowledgement that the candidate has completed them properly. The school examination must be completed and the results submitted to the Inspectorate before the national examination starts (EACEA, 2013d).

The national examination consists of the same questions - or questions of an equivalent degree of difficulty - for all students and is assessed against national standards. National examinations are held in Dutch, English and all other subjects included in the four fixed subject combinations. Courses in subjects chosen as part of the optional component are concluded, in principle, with an examination portfolio.

The national examination consists of tests with open or multiple-choice questions and in some cases a practical component (EACEA, 2013d).

The national examination can be taken at three times during the school year – in May, June and August. All candidates sit the examination in May. The June and August sessions are for students doing resits, or who were unable to sit the examination in May.

The head teacher is responsible for determining each candidate's final marks. The final mark in each subject is the average of the mark for the school examination and the mark for the national examination. To obtain a leaving certificate, a candidate must have scored pass marks in a specified number of subjects. For subjects with only a school examination, the mark obtained is the final mark (rounded off) (EACEA, 2013d).

A scale of 1-10 is used for awarding marks. A mark of 1 is 'very poor', while 10 is given for 'excellent'. A mark of 6 ('satisfactory') constitutes a pass. Students' individual progress is recorded in school reports.

Candidates with a final mark of six or higher in every subject pass their school leaving examination. Students can, however, still be awarded an overall pass mark even if they get a lower mark in some subjects (EACEA, 2013d).



Successful VWO candidates receive a transcript listing the grades achieved in the school examination; the examination syllabus followed for each subject; the grades achieved in the national examination (where applicable); the topic or title of the project undertaken together with the subjects studied for it and the mark obtained; the mark obtained for their specialist subjects; the final grades obtained for the examination subjects; and the overall outcome of the school leaving examination. All successful candidates also receive a certificate listing all the subjects which contributed to the outcome of the examination (O'Donnell *et al*, 2012)

What methods are used to establish standards (for each relevant qualification)

National examinations

All written, national final examinations are set and taken under the supervision of the Government. The results of these written examinations are weighted nationally, so that it makes no difference where a student takes his/her examination. This also guarantees the equivalence of every diploma attained through this system.

Development of national examinations

The Ministry of Education determines the goals for each subject in the examination programme. These are available from the Dutch examinations [website](#). The content of examinations is regularly reviewed to ensure that it is current and reflects developments in society and science. Review committees are chaired by a respected independent chairman; the secretariat is staffed by the SLO (the National Institute for Curriculum Development); and committee representatives are drawn from the trade unions and CITO (the National Institute for Educational Measurement). There are also field consultations during which feedback is sought from stakeholders, such as teachers and textbook authors.

In addition, an independent administrative body, set up by the Ministry of Education (the *College voor Examens*, College Board, is responsible for determining the exact content of the central national examinations and how they are assessed. This is based on a framework contract specifying the number of items to be included in the examinations, the type of items to be included (open, closed, multiple choice) and the aids permitted for students (e.g. calculators). Individual test papers are developed by a team of about three subject teachers under the guidance of an experienced expert CITO colleague.

It takes about two years to develop a new examination paper. The first year is for the development and setting of the tasks/test items. In the second year, examinations are made print-ready, printed and distributed to schools. (Collge voor Examens, 2013)

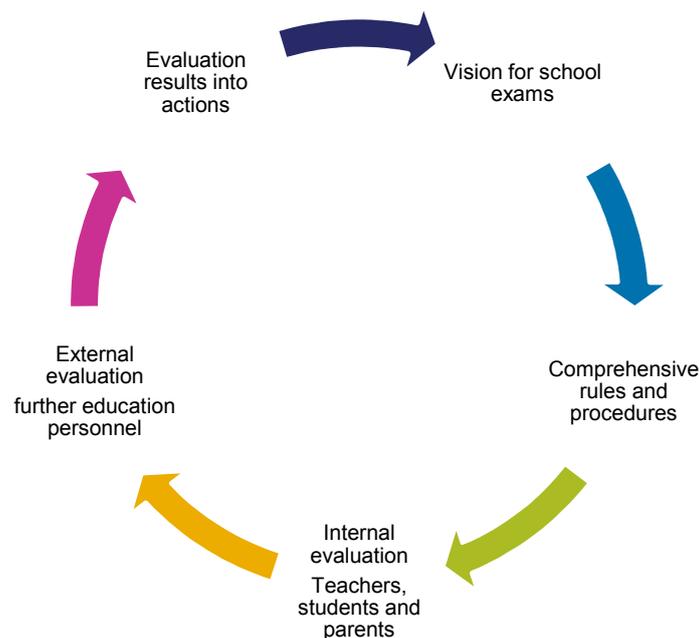
Development of school examinations

The 'Quality Assurance Standard for School Examinations' provides guidance for accountability on the content and processes of the school examinations. This aims to



help schools to standardise the school examination. It was developed in collaboration with school leaders, administrators, the Ministry of Education, Culture and Science and the Education Inspectorate (VO Raad, n.d.).

The introduction to the guide states that rigorous standards for school examinations are needed in a climate where each year there is controversy about the quality of VWO school examinations with suggestions that they are too easy and standards too low. It also says that schools must be accountable for all aspects of the design of school examinations. The mechanism for ensuring high quality school examinations is as follows: (VO Raad, n.d.)



There is also a CITO assessment framework for the quality of school examinations - (Schoolexamensvo.nl, 2012). Unfortunately, this is not available in English.

What methods are used to maintain standards over time (for each relevant qualification)?

The *College voor Examens* (the College Board), set up by the Ministry of Education in 2009, is the independent administrative body for national examinations. It is responsible for ensuring the quality of national examinations in the Netherlands and for ensuring a gradual decrease in the number of examinations taken by students. The Board is aware of the social and political need for a reliable qualification with consistent standards at the end of the different educational pathways and this is a key element of its work (College voor Examens, 2013).

The Board is responsible for standards setting in national examinations, whereas the National Institute for Educational Measurement (CITO) is responsible for overseeing assessment and developing tests, assessment systems and examinations for the whole of the education system. At the request of the Minister of Education, CITO provides an examinations service, covering all aspects from development through to



implementation and invigilation (EACEA, 2013d).

Following an audit of the standardisation system in 2009, a paper was produced describing the standardisation process. This is available [online](#).

Do monitoring assessments exist?

For each national examination, the target performance requirement is exemplified by the development of a reference test and a reference standard for each examination. A reference examination is a national examination from a previous final year that is widely regarded as a good representation/operationalisation of substantive examination requirements (College voor Examens and CITO, 2011).

What methods are used to review standards, e.g. are there post-examination reviews, and how are the results used?

When and how are the results reported and to whom?

There have been system audits, the results of which have been fed back into improving the system.

In addition, CITO collects teacher judgements on the examinations through brief questionnaire surveys. In some cases, this is a 'quick scan' survey questionnaire followed by a digital questionnaire. This information is primarily intended for the further development of examinations and not to support the maintenance of standards (College voor Examens and CITO, 2011).



New Zealand

Period of compulsory education

Education, through full-time attendance at school, is compulsory for everyone between their sixth and sixteenth birthday. Students may enrol and receive a free education at any state school from their fifth birthday until first January following their nineteenth birthday (or their twenty-first birthday for students with special educational needs) (O'Donnell *et al.*, 2012).

Secondary education caters for students aged 13 to 18 years, although compulsory lower secondary schooling ends at age 16. Most secondary schools are Year 9 to 13 (Form 3-7) schools (students aged 13-18), although in rural areas there are Year 7 to 13 (Form 1-7) schools (age 11-18) and composite schools ('area schools') catering for the full age range, five/six-18 years (O'Donnell *et al.*, 2012).

What high stakes school leaving academic qualifications exist and when are they taken?

The National Certificate of Educational Achievement (NCEA) is the main qualification at all levels of the senior secondary school and is the high stakes academic qualification. Its purpose is to provide opportunities for the diverse range of students in the increasingly wide variety of courses in schools to have their achievements recognised and reported. Consequently, students can achieve the NCEA from a wide range of studies, within and beyond the school curriculum. Unit and achievement standards are used for assessment of courses developed by education, industry and national standards bodies. For each school curriculum subject, there is both external assessment and externally moderated, internal assessment, using achievement standards. Achievement standards are in place for general/academic (school curriculum) subjects, such as maths, geography and science) and unit standards have been developed for assessment in vocational/technical subjects (e.g. automotive engineering, dairy manufacturing).

The NCEA is a qualification on New Zealand's Qualifications Framework (NZQF); it sits alongside 700 other national qualifications used throughout tertiary education and industry training. The NZQF was developed by the New Zealand Qualifications Authority (NZQA) in response to the Government's belief that:

the social and economic needs of a modern society require a flexible and cumulative qualifications framework that recognises achievement over a diverse range of subjects and levels, from practical to academic, and acknowledges that education and training can take place in different institutions and at different times in a person's life.

(New Zealand, Ministry of Education, 1997)

Credits are gained when students achieve 'standards' which specify what a person knows and can do in a particular area of learning (either a conventional school subject or a vocational or technical subject). A NCEA is attained when a specified



number of credits at particular levels of the New Zealand Qualifications Framework has been gained. Some students may complete a NCEA by the end of secondary school. Others may carry the credits they have achieved at school through to post-school education and training and complete their NCEA, a national diploma, or degree at the tertiary level. All credits towards national qualifications are recorded on a national database maintained by the New Zealand Qualifications Authority (NZQA). The aim is to create a 'seamless' education system, based on unit standards, which integrates traditional senior secondary education, industry training and tertiary education. NCEAs are available at levels 1, 2 and 3. Level 1 units are generally taken in the first year of upper secondary education (O'Donnell *et al.*, 2012)

Further details on the National Qualifications Framework are available at <http://www.nzqa.govt.nz/studying-in-new-zealand/nzqf/>

Scholarship examinations

Scholarship provides recognition and monetary reward to top students in their last year of schooling. Students continue to receive their award as long as they maintain a B average. Scholarship examinations enable candidates to be assessed against challenging standards and are demanding for the most able candidates in each subject. Scholarship candidates are expected to demonstrate high-level critical thinking, abstraction and generalisation, and to integrate, synthesise and apply knowledge, skills, understanding and ideas to complex situations. Approximately three per cent of Year 13 students studying each subject at level 3 are awarded scholarship if they reach the standard that has been set. There is some tolerance around the three per cent, mainly for those subjects with few candidates. Students can enter for scholarship through an accredited New Zealand secondary school (NZQA, 2013d).

University entrance

To gain entry to a New Zealand university, students must have 42 credits at level 3 or above on the New Zealand Qualifications Framework (NZQA, 2013a).

Universities and other tertiary providers in New Zealand also sometimes have other specific requirements for entry to particular programmes or courses (NZQA, 2013a).

What format do they take, how does testing take place and what is tested?

National Certificate of Educational Achievement

The NCEA (National Certificate of Educational Achievement) is a composite qualification made up of credits gained from a wide range of subjects including traditional school curriculum areas and alternative programmes that are listed on the New Zealand Qualification Framework (NZQF). NCEAs are available at levels 1, 2 and 3. Level 1 units are generally taken in the first year of upper secondary education (O'Donnell *et al.*, 2012).

Students start to obtain credits (on the NZQF) towards the NCEA in Year 11, the final year of compulsory education, aged 15/16. A student's achievement of the



standards required to obtain credit towards the NCEA is measured by an appropriate mix of external and moderated internal assessment. At least half of the credits for each subject are assessed externally. External assessment includes written examinations and other assessment such as portfolios of students' work (O'Donnell *et al.*, 2012).

The NCEA assessments are designed to suit the skill or knowledge being tested. In school curriculum subjects, at least half of the assessment is based on end of year examinations run by the New Zealand Qualifications Authority (NZQA). Schools can also use other standards on the National Qualifications Framework to tailor courses for the particular needs and requirements of students. In this way, they can prepare all students for careers in areas of their choice (O'Donnell *et al.*, 2012).

Examinations and assessments for the purpose of awarding secondary school qualifications are the responsibility of the New Zealand Qualifications Authority (NZQA). Standards for these examinations and assessments are based on the learning outcomes of the national curriculum statements (O'Donnell *et al.*, 2012).

A minimum number of credits is required, from examinations and internal assessments, for students to achieve the award level for a National Certificate at any level. Students can accumulate credit towards their certificate over a number of years. They do not, for example, have to achieve all the credits for a level 1 certificate while in Year 11 (the final year of compulsory secondary education, age 15-16), although it is expected that reasonably able students should be able to do so.

- For students to be awarded NCEA level 1 they must achieve 80 credits. Of these eight must be in literacy standards and a further eight must be in numeracy standards.
- For students to be awarded NCEA level 2, they must achieve 80 credits, 60 of which must be from level 2 standards.
- For NCEA level 3, students must achieve 80 credits, 60 of which must be from level 3 standards, the remaining 20 from level 2 (O'Donnell *et al.*, 2012).

In each area of learning, different aspects of skills, knowledge and understanding are assessed separately. Each aspect can earn a different number of credits.

Consequently, results about each student's efforts are detailed, giving a profile of their achievements. The profile shows a student's credits and grades for each national standard. Results also show how each student performed compared with all other students across the country. Finally, they provide a grade average for each subject or area of learning (O'Donnell *et al.*, 2012).

When students achieve a standard for an area of learning, they earn credits towards their NCEA qualification (at one of the levels detailed above). In school curriculum subjects, students can do more than achieve a standard - they can achieve with merit or excellence grades (O'Donnell *et al.*, 2012).

Scholarship examinations

New Zealand scholarship standards have been aligned with the New Zealand Curriculum (NZC). Examinations have between three and six questions and the



scores for each question are indicated in the candidates' examination booklets, which are returned soon after the results (NZQA, 2013d).

All examinations last for three hours. They are taken in November (the New Zealand school year starts in January). The Scholarship results are released towards the end of February each year (NZQA, 2013d).

Markers allocate a score of 0-8 for each question in the scholarship examinations using subject specific schedules. These are based on a generic marking guide, as follows:

- An answer given a score of 8 is an outstanding answer in all respects. Strong evidence of integration and synthesis. As good as could be expected under examination conditions. Accurate, comprehensive, coherent, lucid, perceptive.
- A score of 1 shows meagre understanding relevant to the question.
- 0 is awarded for answers that are blank or irrelevant (NZQA, 2013d).

The work of the highest scoring candidates is reviewed by marking panels to identify the top candidates in each subject. Examination booklets for candidates whose scores are close to either the scholarship or outstanding cut-offs are also re-marked, to ensure their results are correct (NZQA, 2013d).

A selection of examination papers has been reproduced with permission from previous 'top scholars'. To view these examination papers, go to the [scholarship subjects](#) page and click on the 'assessment specification and resources' link for each subject. The top scholar examination booklet is at the bottom of the page under the [subject] resources heading.

What methods are used to establish standards (for each relevant qualification)?

NZQA has a formal quality assurance process to ensure that the assessment of each standard is fair across all students, regardless of the school they attend. This includes internal moderation, external moderation and assessment system checks.

The [Directory of Assessment Standards](#) (DAS) lists all quality assured unit and achievement standards, known collectively as 'assessment standards'. The assessment standards listed on the DAS can contribute to all national qualifications, that is, they are valid for both NCEA and scholarship assessments and examinations.

Unit standards are developed by industry training organisations and by two NZQA units – National Qualifications Services and Māori Qualifications Services. The Ministry of Education is the only developer of achievement standards, which are derived from the achievement objectives of the New Zealand Curriculum.

Collectively, the developers of assessment standards are known as standard-setting bodies (SSBs) (NZQA, 2013c).

The use of DAS assessment standards in nationally recognised qualifications helps to ensure that:

- clear outcomes are recognised



- nationally consistent standards apply to the outcomes that are recognised.
- existing knowledge and skills are recognised and credited on the Record of Achievement.

The DAS is supported by a three-pronged quality system:

- Quality assurance of standards – National assessment standards are quality assured before being listed on the DAS.
- Consent to assess against standards (until recently referred to as NQF accreditation) – Organisations demonstrate that they are able to develop or access assessment resources, undertake internal moderation, engage in the SSB's national external moderation system, and report results in a timely manner.
- National external moderation of assessment – The developer SSB runs a moderation system that ensures national consistency of assessment decisions against DAS standards (NZQA, 2013c).

Each standard listed describes what a candidate who has achieved the standard knows and can do. Each standard has a defined credit value, which represents the notional learning time, and a level, which reflects the level of complexity of the skills and knowledge that are recognised by the standard. The common currencies of credit values and levels enables the credits gained from standards to be portable among national qualifications.

What methods are used to maintain standards over time (for each relevant qualification)?

This FAQ from the NZQA website details how standards are maintained across New Zealand (NZQA, 2012b):

New Zealand is an optimum size to achieve nationally consistent internal assessment. We are large enough to have a complete system and a very sound pool of expertise – but we are small enough for all teachers in a subject to be in touch with the national system. We have a professional community of understanding about NCEA standards and processes.

Compared with teachers in many other countries, New Zealand teachers are assessment experts. A large proportion of teachers have had experience in setting and marking examinations, in writing national standards, and as moderators or as members of moderation clusters.

An important part of the quality assurance system for NCEA is called external moderation – making sure teachers are making consistent internal assessment decisions across the country. The aim is to equip teachers to make accurate and consistent judgements, by providing feedback and professional development.

How are teachers' assessment decisions checked for national consistency?

NZQA uses a number of methods to monitor the consistency and accuracy of internal assessment:



- NZQA employs over 34 full-time equivalent moderators and 235 part-time moderators. Most moderators are current or recent teachers and all are assessment experts in particular subjects.
 - Moderators run best practice workshops, develop resources to guide schools and speak to meetings of subject associations. Moderators also check each school's assessment tasks and activities, and the judgements schools are making when they assess student work.
 - Moderators currently check a sample of about ten per cent of each school's internally assessed work. This is ample to show whether or not teacher judgements are consistent across the jurisdiction.
 - NZQA calculates agreement rates - measures of the extent to which moderators and teachers agree on whether samples of student work meet the standards. Each school receives a report on the quality of its internal assessment. These are called Managing National Assessment reports and they are [published on the NZQA website](#). Moderation reports are quite specific. For example, for a particular achievement standard a school could be making entirely accurate decisions about Achievement and Merit, but getting it wrong in awarding Excellence.
 - NZQA compares each school's results from internally-assessed and externally-assessed standards in each subject. Internal and external achievement rates differ nationally and it is expected that each school will broadly reflect national patterns. If a school's internal results are greatly different from what is expected on the basis of their external results (and if the teacher-moderator agreement rates are poor) NZQA works with the school to improve its internal assessment processes.
 - As an ultimate sanction, NZQA can withdraw a school's right to assess for national qualifications; this is most likely to apply to particular subjects within a school. In that case, provisions would be made for students to be assessed through another school's quality systems.
 - Does the work meet the level described for Achievement? Should the student gain credits for that standard?
 - Has the student done well enough to get a grade beyond Achievement? If so, will it be Merit or Excellence?
 - NZQA calculates two agreement rates - how well moderators and teachers agree on awarding credit for the standard, and how well they agree on the specific grade for that standard.
 - Agreement rates for awarding credit are always higher than agreement rates for grades, as statistically, fewer decisions are involved in deciding on credit.
- In many standards there's a fine distinction between the grades of Achievement and Merit, and between Merit and Excellence. In 2010, 97 per cent of assessment materials were deemed to be suitable, either unmodified or with only minor modification. This high figure reflects the fact that most schools now use



downloaded assessment materials that have been pre-approved.

Moderation is a professional interaction

Moderators run assessment workshops for teachers in individual subjects, all around the jurisdiction [...]

Moderators work hard to provide teachers with ‘clarification statements’ where there appears to be some difficulty or confusion about some aspect of the required standard.

Schools can ask for clarification or appeal a moderator’s decision. In 2009 and 2010, fewer than one in 1,000 moderator judgements were successfully appealed.

In many regions, schools voluntarily form clusters to enable teachers to compare notes with others teaching their subject. This is especially valuable where there are only one or two teachers of a particular subject in a school. In addition to enhancing NCEA assessment, clusters provide professional development for teachers.

Agreement rates

For each achievement standard, students receive a set number of credits if their work reaches the Achievement level. If they exceed the Achievement level, they can gain Merit or Excellence grades. So in effect, teachers make two decisions when they assess student work:

In 2010, across all standards at all levels, moderators agreed with 91 per cent of teachers’ assessment judgements in awarding credit. This was up from 83 per cent in 2009, and was higher than in any previous year.

2010 moderator/teacher agreement rates across all subjects:

	Level 1	Level 2	Level 3	Across all levels
<i>Agreement on credit</i>	94%	93%	91%	91%
<i>Agreement on grades</i>	86%	85%	82%	84%

Do monitoring assessments exist?

Monitoring assessments only exist at the primary level, where the ‘National Monitoring Study of Student Achievement’ provides some assessment of ability in mathematics (plus health and physical fitness).

The New Zealand Council for Educational Research also provides some tests to help teachers confirm assessments but these are not compulsory. Further details are [available](#).



What methods are used to review standards, e.g. are there post-examination reviews, and how are the results used?

Standards and national qualifications are reviewed on a regular basis. They may also be revised if changes need to be made outside of the formal review process (NZQA, 2012a).

In a revision, changes are made to unit standards or national qualifications that maintain the achievability of the standard or qualification until it is next reviewed. Evidence of appropriate consultation is provided in support of the proposed changes. A revised standard or national qualification is published as a new version, but retains its original registration date (NZQA, 2012a).

A review is a planned process that involves consultation with all stakeholder groups, and results in a new version of a standard or national qualification, which is deemed fit for purpose. Reviewed standards and national qualifications are normally re-registered for a period of five years (NZQA, 2012a).

New unit standards are first registered for a period of approximately five years. When a standard reaches its planned review date and there is insufficient information to inform a review, it may be rolled over for a period of time, consistent with the SSB's review schedule. Before rollover, the standard is checked to ensure that all references, including legislation and prerequisites, are current. If standards require amendments prior to a full review, as well as an extension of their period of registration, the process is called 'revision and rollover'. The review considers whether the standard:

- meets current quality assurance criteria
- is fit for purpose for another five years
- takes into account any updated requirements in the industry, professional or academic area to which it relates
- has any shortcomings, identified through assessment and moderation (NZQA, 2012a).

Qualifications may be revised to take account of new factors or omissions; changes to legislation; attainability issues; or a review of standards contained in the qualification.

A revised qualification is published as a new version, but retains its original registration date. A change report is published on the NZQA website, covering the consultation process and the changes made to the qualification (NZQA, 2012a).



Singapore

Period of compulsory education

In Singapore, there is an average of ten years of formal general education, comprising six years of compulsory primary school education and four years or more in the secondary sector. Secondary education is not compulsory although attendance is generally universal (O'Donnell *et al.*, 2012).

Primary education commences in the January of the year in which a child reaches the age of seven; for example, a child whose date of birth fell between 2nd January 1996 and 1st January 1997, would have begun primary education on 1st January 2003 (O'Donnell *et al.*, 2012).

Upper secondary education, known as post-secondary (or pre-university) education in Singapore, caters for students aged 16/17 to 18/19/20, in three different types of institution:

- Junior colleges, which offer two-year pre-university courses leading to the Singapore Cambridge (International) General Certificate of Education Advanced Level (GCE 'A' Level) examinations and tertiary education.
- Centralised institutes, which offer three-year pre-university courses leading to the Singapore Cambridge (International) GCE 'A' Level and tertiary education.
- Polytechnics, which offer three-year courses leading to a diploma (O'Donnell *et al.*, 2012).

What high stakes school leaving academic qualifications exist and when are they taken?

There are a series of high stakes examinations in Singapore. Children take the national Primary School Leaving Examination (PSLE) at the end of Primary 6 (age 12). Their results in this examination determine their placement on differentiated courses of lower secondary education (e.g. normal academic, normal technical, special/express), according to their learning pace, ability and inclinations (Ministry of Education, Singapore, 2012).

The Singapore-Cambridge GCE 'N' Level and GCE 'O' Level examinations are national examinations which measure student attainment on completion of lower secondary (known as 'secondary') education (aged around 16).

Students with good GCE 'O' Level passes are normally admitted to junior colleges, where they complete the Singapore-Cambridge GCE 'A' Level in two years, or to centralised institutes to complete the 'A' Level in three years. Both GCE 'O' Level and GCE 'A' Level students can apply for entry to polytechnics. Students who have completed their GCE 'N' Level or GCE 'O' Level examinations may also apply for various technical or business study courses (O'Donnell *et al.*, 2012).

Student attainment in the GCE 'O' Level and GCE 'N' Level examinations is recorded as Grades 1, 2, 3, 4, 5 and U (ungraded). Performance in the oral/aural examination



in Chinese, Malay and Tamil is indicated as Distinction, Merit or Pass (O'Donnell *et al.*, 2012).

Students who have completed secondary school education and taken the GCE 'N' Level or GCE 'O' Level examinations, but who do not qualify for the next higher level examination, usually seek employment (O'Donnell *et al.*, 2012).

Since 2008, all students who complete their education at the secondary and pre-university levels have also received a comprehensive school testimonial called the School Graduation Certificate. This is provided by the Ministry of Education (MOE) and includes academic and non-academic achievements and personal qualities (O'Donnell *et al.*, 2012).

With the implementation of the 2006 GCE 'A' Level curriculum students have had access to a broader, more multi disciplinary learning curriculum. They now choose from a set combination of subjects as follows (Ministry of Education, Singapore, 2007a):

Three H2 content-based subjects	at least one of these must be from a contrasting discipline
One H1 content-based subject	
One H1 mother tongue language	
H1 general paper	
H1 project work	

H2 subjects are equivalent to previous 'A' level subjects in terms of demand and intellectual challenge. Most H2 subjects have similar coverage to previous 'A' level subjects but content is reduced to free up curriculum time for contrasting subjects and non-academic pursuits. An H1 subject is equivalent to half of an H2 subject in terms of curriculum time but it is of the same intellectual level. H3 programmes are meant to allow exceptional students to pursue a subject or area in which they have the interest and aptitude (Ministry of Education, Singapore, 2007b).

Students may also offer 'Knowledge and Inquiry' (KI¹³) in place of the general paper, or offer mother tongue language and literature at H2 level (Ministry of Education, Singapore, 2007a).

¹³ Knowledge and Inquiry has both a knowledge (content) component and a skills-based component. It may be offered in lieu of a general paper (GP). Students offering KI must offer it as their fourth H2 subject. KI is an intellectually demanding subject. Students offering the subject should have a good command of the English Language, a curious mind and a keen interest in the world around them. They are expected to do much reading. There is an independent study component where students are required to work on a research topic of their choice for six months, so students offering KI should be capable of and comfortable with working independently.



The grading system for the three levels of study is shown below: (Ministry of Education, Singapore, 2007a):

H1 Subjects	H2 Subjects	H3 Subjects
A	A	Distinction
B	B	Merit
C	C	Pass
D	D	Ungraded
E	E	
S	S	
Ungraded	Ungraded	

'S' and 'Ungraded' are grade points below pass; 'S' is used to denote a sub-pass

In an academic year which ends in March/April, students sit the examinations for the GCE normal technical and normal academic ('N' level) qualifications between July and October, for the GCE 'O' Level between July and November and, for the GCE 'A' Level between July and December. The timetables indicate that oral examinations take place at the beginning of the examination session with written examinations at the end (SEAB, 2013).

Examination results for the GCE normal technical and normal academic qualifications are released in December of the examination year, in the January of the following year for GCE 'O' Levels, and in March of the following year for GCE 'A' Levels (SEAB, 2013).

What format do they take, how does testing take place and what is tested?

The Singapore-Cambridge GCE 'N' Level and GCE 'O' Level examinations are national examinations which measure student attainment on completion of lower secondary (known as 'secondary') education (O'Donnell *et al.*, 2012).

Students on the special/express course usually take the Singapore-Cambridge GCE 'O' Level examination in seven to nine individual subjects at the end of secondary year four (S4, age 16). Students in the normal academic and normal technical courses take the Singapore-Cambridge GCE 'N' Level examination at the end of S4. Students on the normal academic course usually take the examination in six to eight individual subjects; those on the normal technical course in five to seven subjects (O'Donnell *et al.*, 2012).

In the above GCE examinations, candidates may take written, oral and/or practical



examinations, depending on the subjects they are studying. Candidates are also assessed through coursework in subjects such as design and technology, food and nutrition, and fashion and fabrics (O'Donnell *et al.*, 2012).

For most subjects, the relationship between curriculum and assessment is incorporated in the syllabus documents. For some subjects, e.g. English language, the teaching syllabus includes suggestions for informal and diagnostic assessment, while a separate examination syllabus specifies the skills to be examined and the allocation of marks for each skill. Syllabus documents are available online:

[GCE A levels](#)

[GCE O levels](#)

[GCE N \(T\) \(technical\)- Level](#)

[GCE N\(A\) \(academic\) -Level](#)

What methods are used to establish standards (for each relevant qualification)?

The Singapore Examinations and Assessment Board (SEAB) was established on 1st April 2004 as a statutory board. SEAB, formerly the Examinations Division of the Ministry of Education (MOE), was formed to develop and conduct national examinations in Singapore, and to provide other examination and assessment services and products, locally as well as overseas. SEAB collaborates with MOE on all national examinations.

Since 2006, MOE has taken greater responsibility for developing examination syllabuses and formats, setting standards and awarding grades. Despite this, MOE and the Singapore Examinations and Assessment Board (SEAB) continue to work with University of Cambridge International Examinations in designing syllabuses, and outsource to UCLES the setting of question papers and marking of examination scripts (Ministry of Education, Singapore, 2013).

There is no change to present procedures as far as national examinations are concerned. SEAB will continue to develop reliable and valid instruments and ensure that the award of qualifications is based on accurate assessment and rigorous standards. (Ministry of Education, Singapore, 2013).

What methods are used to maintain standards over time (for each relevant qualification)?

Do monitoring assessments exist?



What methods are used to review standards, e.g. are there post-examination reviews, and how are the results used?

--



South Korea

Period of compulsory education

The Education Principle Law stipulates that everyone has the right, as well as the duty, to be educated as a citizen. Every Korean has a right to nine years of education, comprising six years of compulsory elementary school education and three years of lower secondary education in junior high school. Education is therefore compulsory between age six and age 15 (O'Donnell *et al.*, 2012).

Post-compulsory upper secondary (high school) education (students aged 15-18) aims to provide advanced general and specific education, building on lower secondary (junior high school/middle school) education. High schools are classified into general/academic, vocational/technical and other specialist/special purpose high schools (for talented/gifted students) - art/music and athletic, foreign language and science schools. Students therefore need to select the course they wish to follow. The courses last three years and students bear the costs of education. Students completing their studies in high school are awarded a high school diploma (O'Donnell *et al.*, 2012).

What high stakes school leaving academic qualifications exist and when are they taken?

On completion of upper secondary high school education, successful students receive the diploma of high school education.

There are also other high stakes assessments during this phase of education in Korea, they include:

- entrance examinations to high schools (although these are diminishing)
- aptitude tests on completion of this phase for entry to higher education (College Scholastic Ability Tests, CSATs)
- the nationwide assessment system/national assessment of educational achievement (annual national scholastic achievement tests (SATs), for different year groups and different subjects each year)
- continuous classroom assessment by teachers which monitors student achievement - often by monthly testing - the results of which are included in the record of achievement, which accompanies the high school diploma (O'Donnell *et al.*, 2012).

Admissions to higher education are based on students' high school records, their involvement in extra-mural activities and their scores in the College Scholastic Ability Tests (CSATs). Korean universities also consider students' special talents or aptitudes, and their social background in determining admission (special consideration is given to those who are rural residents and/or socially disadvantaged, for example). However, among all factors, CSAT scores are the most important in making admissions decisions (KICE, 2012).



The level of national attention afforded to the CSAT is so high that the test questions are revealed and evaluated by experts in newspapers, on TV, and on radio immediately after completion of the test (KICE, 2012).

Junior college entrance is also dependent on school achievement, for example, results in the CSAT, along with an interview and aptitude test. (O'Donnell *et al.*, 2012)

The College Scholastic Ability Tests (CSATs) for higher education entry take place once a year, usually in mid November. The CSATs are one-day tests which begin at 8:40 am and end at 17:30. In the CSAT, a total of 230 test items are taken in 400 minutes, for a maximum score of 400. Most of the test items are five-option multiple choice questions with some short answer items (O'Donnell *et al.*, 2012).

CSAT results are released to students in mid-December in a academic year which runs from March to December (Card, 2005).

For the purposes of this study, we focus on the CSAT as the high stakes general/academic examination.

What format do they take, how does testing take place and what is tested

High school certificate/diploma

The high school certificate/diploma is awarded once a student has successfully completed 215 units of study. Each area of study is allocated a certain number of units, with one unit being equal to 50 minutes. Studies at this level are split into three categories:

- National Common Basic Curriculum (NCBC) subjects which include:
 - Korean language
 - moral education
 - social studies
 - mathematics
 - science
 - practical arts
 - physical education
 - music
 - fine arts
 - foreign language (English)
 - optional activities
 - extracurricular activities
- General subjects
- Specialist subjects

The general and specialist subjects expand upon the NCBC core areas and can also include further foreign languages, Chinese characters, the classics and liberal arts (O'Donnell *et al.*, 2012).

Although Grade 10 still consists of the subjects from the National Basic Curriculum,



as of Grade 11, students can choose between three subject clusters: humanities, sciences, or vocational training (O'Donnell *et al.*, 2012).

College Scholastic Ability Test (CSAT)

CSAT has two objectives. First, to provide fair and objective information in the selection process for university on the basis of a valid and reliable test; and second, to 'normalise' high school education by developing tests based on the content and level of the high school curriculum (KICE, 2013).

CSATs are norm-referenced and the results are recorded by numerical scores (O'Donnell *et al.*, 2012).

Since 2005, CSAT has also involved a written test. Students choose five of the subjects from the list below, all of which focus on high-order thinking abilities, analytical ability and the student's understanding of cross-disciplinary materials. Their choice depends on the universities and colleges they are applying to. The language section, for example, assesses a student's ability to use Korean appropriately and effectively for study at college or university and tests listening, speaking, reading and writing ability with integrated subject materials. It also evaluates vocabulary, factual comprehension, inferential/imaginative comprehension, critical comprehension and logical comprehension ability. The results are expected to reflect the student's ability to use vocabulary and to think analytically, logically and creatively. The tests are available in the following subject areas:

- Korean language
- Mathematics: a choice of one from differentiation and integration, probability and statistics, or discrete mathematics
- English: including listening and speaking items
- Social studies, science, and vocational education: from each of these areas, students select a topic on which to be tested, as follows:
 - Social studies: students select from a choice of four which includes Korean, geography, world geography, economic geography, Korean modern and contemporary history, Korean history, world history, law and society, politics, economics, society and culture, ethics
 - Science: a choice of four from physics I, physics II, chemistry I, chemistry II, biology I, biology II, earth science I, earth science II
 - Vocational studies: a choice of one from agriculture, industry, commerce, fishery and marine transportation, home economics and vocational education
- A second foreign language from German, French, Spanish, Chinese, Japanese, Russian, Arabic or Old Korean (O'Donnell *et al.*, 2012).

Item weights vary depending on the importance of content and the required time. Multiple choice items have five choices. The (first) foreign language section of the CSAT assesses a student's ability to use English fluently. Test items include the four language skills of listening, speaking, reading and writing (O'Donnell *et al.*, 2012).

The gradual addition of various elective subjects to the CSAT is a relatively recent development and reflects the trend in high school education for common compulsory



subjects to be reduced, whilst elective ones are expanded within the curriculum (O'Donnell *et al.*, 2012).

The revised (2005) CSAT is based on material studied during Years 11 and 12 of education (students aged 16-18). Previously, 70 per cent of the material for testing was taken from Year 10 (the first year of post-compulsory upper secondary education for 15-16 year olds) and only 30 per cent from the final two years of post-compulsory education (O'Donnell *et al.*, 2012).

On the completion of CSAT, answer sheets are sent from the test sites across the jurisdiction to the Korea Institute for Curriculum and Evaluation (KICE) in Seoul for marking. Marking takes approximately three weeks. During the process, the answers are checked for errors. The answer sheets are then marked on the basis of the correct answers and scores are calculated according to the designated scoring methods. When the CSAT scores are notified to individual students, KICE releases the results of that year's grading through the media. This includes the number of applicants, score characteristics by different applicants, scores by subject/area, stanines, etc. (KICE, 2013).

What methods are used to establish standards (for each relevant qualification)?

College Scholastic Ability Test (CSAT)

The Korea Institute of Curriculum and Evaluation (KICE) is responsible for the College Scholastic Ability Test (CSAT). KICE's responsibility for the CSAT involves determining the basic action plan for the CSAT, developing the test items, and analysing and reporting the test results. Before starting to develop test items, the CSAT Administration Committee holds several meetings to obtain advice on developing test items, recruiting test developers and providing workshops for designers. Major activities are as follows:

- An Advisory Committee provides advice on selecting item developers and item reviewers, on item development techniques and on training new item developers
- Once item developers are selected, they are isolated in a secure place for a period of time to develop test items. Developers and reviewers include university professors and high school teachers
- Examination papers are printed under heavy supervision in a secure place and are then distributed country-wide. Manager-level government officials from the metropolitan and provincial education authorities (MPEAs) are responsible for the distribution of CSAT papers in their area (O'Donnell *et al.*, 2012).

A strict administration schedule for CSAT assessment is provided to schools by KICE. This determines, for example, the exact time students are expected to enter the designated testing area, in addition to the exact times when teachers are expected to administer specific individual tests (O'Donnell *et al.*, 2012).

Scoring of the CSAT tests - by OMR (optical mark recognition) answer sheets - is



carried out under the supervision of KICE (O'Donnell *et al.*, 2012).

Results are distributed to the metropolitan and provincial education authorities (MPEAs) as well as to individual students. Since the 2008 academic year only stanine scores (grade scores) have been reported. The stanine scores are single-digit scores ranging from one to nine, in which one is high and nine is low. They are obtained by dividing the raw score into nine parts (O'Donnell *et al.*, 2012).

The following measures aim to ensure the fairness of CSAT. First, the entire process of selecting test organisers, developing items, editing, and printing test booklets is controlled in strict security. The team of test organisers, who are selected under strict criteria, pledges to develop items with impartiality and fairness and keep them secret. They ensure the objectivity and validity of the test through various stages of examination and development meetings. In this process, special efforts are made to exclude items already used in study-aid books and workbooks for private education because they are considered advantageous to those students who have studied them. All those who participate in the process of editing and printing test items also take the same responsibility (KICE, 2013).

In the process of implementing CSAT, those who transport test booklets, manage test sites, and administer tests are given training in advance and, along with those completing the CSAT, they are isolated from the outside world throughout the testing time. In particular, cases of cheating on the tests are sternly punished by prohibiting the offenders from applying for the CSAT for the next three years (KICE, 2013).

During the process of marking, the answers are checked for errors. In this process, the question of whether the content of the test is consistent with the high school curriculum or advantageous to particular test-takers is discussed again (KICE, 2013).

What methods are used to maintain standards over time (for each relevant qualification)?

College Scholastic Ability Test (CSAT)

The CSAT is a norm referenced test – graded on a curve (KICE, 2013).

There are standard scores for each test of Korean language, mathematics and foreign language (English). They have a mean of 100 and a standard deviation of 20 (KICE, 2013).

The standard scores for each subject in the social studies/science/vocational education and second foreign languages/Chinese characters and classics tests have a mean of 50 and a standard deviation of 10 (KICE, 2013).

The table below shows the standard scores conversion of raw scores:

Test	Number of Items	Maximum Raw Score	Standard Scores		
			Mean	Standard deviation	Range



Korean Language	50	100	100	20	0~200
Mathematics	30	100	100	20	0~200
Foreign Language (English)	50	100	100	20	0~200
Social Studies/Sciences/Vocational Education	20	50	50	10	0~100
Second Foreign Languages/Chinese Characters and Classics*	30	50	50	10	0~100

The numbers represent each subject's number of items, maximum raw score, and standard score.

Range = (test/subject mean) \pm 5 X (test/subject standard deviation). (KICE, 2013)

CSAT score reports show the standard score, percentile rank, and stanine for each test/subject. The standard score is calculated through the linear transformation method. Scores among the elective subjects in Math A are adjusted. Standard scores and percentile ranks are rounded-off to an integer. The percentile indicates the percentage of students falling below the midpoint of the given score interval. Stanines are single digit scores ranging from 1 to 9. The distribution of the standard score is divided into nine parts (standard nine) where a stanine score of 1 is the highest (KICE, 2013).

A research abstract on the KICE website discusses the development of an item bank for the CSAT. It states:

The CSAT test items have been developed so far by a group of teachers at universities and high schools who have been confined to a secure place, isolated from the outside world for about a month just before the administration of the test. There have been problems with this method of preparing the test. The whole complex process of developing and refining the test items, printing and distributing the test papers to the whole nation should be completed within less than a month.

In view of the problems this study aims to provide preparatory research for the establishment of an item bank for CSAT which would guarantee provision of high quality test items, administrative effectiveness, and flexible administration of the test as occasion demands. In line with this, the study of 2000 has carried out the following research.

First, as a preliminary step to establish the item bank, 130 items for Korean language and 110 items for foreign languages (English) have been developed as a base provided by the research. There have been discussions over the items previously used in respective fields.

Second, the KICE Item Bank Searcher (KIBS), which can store and retrieve effectively all the items developed for the CSAT item bank since 1997, has been devised. The KIBS is expected to enhance systematic management of the item bank and effective utilisation of the items stored in it.

Third, a new plan for developing, refining, and storing the test items, and a security device for the item bank have been developed in order to help the



administration of CSAT.

The adoption of the item bank system for CSAT is bound by government policy. When the Government determines to adopt the item bank system for CSAT, it will be possible for the Korean Language and Foreign Language items prepared since 1997, and stored in KIBS to be put into actual use

(KICE 2000)

Do monitoring assessments exist?

What methods are used to review standards, e.g. are there post-examination reviews, and how are the results used?

When and how are the results reported and to whom?

As the institution responsible for the development and implementation of CSAT, KICE is fully aware of the importance of CSAT and of concerns over its limitations. As a result, since the launch of CSAT, KICE has undertaken research on the development and evaluation of the CSAT and the utilisation of the results from the viewpoints of educational evaluation/assessment, curriculum development and educational policy. In addition, the results of this research are linked to the data of another similar national-level test, that is, the National Assessment of Educational Achievement and, on the basis of these efforts, various research outcomes have been produced concerning the current education system in Korea. Furthermore, by carrying out research on college entrance examinations and national-level evaluation systems of other countries, KICE attempts to improve and develop the CSAT (KICE, 2013).



Appendix B – Reference summaries

Reference	Jurisdiction	Summary	Type and quality of evidence
Alison, J. (n.d.). <i>A Lesson in How not to Make Education Policy Change: New Zealand School Qualifications and the Struggle for Hegemony</i> [online]. Available: http://www.acsa.edu.au/pages/images/judie_alison.pdf [4 October, 2013].	New Zealand	Discussion of political aspects of the history of the qualification reform in new Zealand during the 1990s.	Historical research/ review paper focussed on the political atmosphere surrounding the curriculum changes. No technical information about how standards were set or maintained.
Backes-Gellner, U. and Veen, S. (2008). 'The consequences of central examinations on educational quality standards and labour market outcomes', <i>Oxford Review of Education</i> , 34 , 5, 569–588.	Germany	This investigation shows that the number of students obtaining the <i>Abitur</i> qualification in Germany has risen steadily over the years. This rise has been steepest in <i>Länder</i> where there is no centralised assessment. The paper shows that the <i>Abitur</i> qualification has become de-valued, with the salary premium for employees possessing an <i>Abitur</i> decreasing over the years.	Research article. Of good relevance.
Dufaux, S. (2012). <i>Assessment for Qualification and Certification in Upper Secondary Education: a</i>	Various	This paper looks at the design of assessment for qualification and certification in upper secondary	A review paper that covers a number of topics in depth. Provides good quality evidence,

Reference	Jurisdiction	Summary	Type and quality of evidence
<p><i>Review of Country Practices and Research Evidence</i> (OECD Education Working Papers, No. 83). Paris: OECD Publishing [online]. Available: http://www.oecd-ilibrary.org/docserver/download/5k92zp1cshvb.pdf?expires=1380886666&id=id&accname=guest&checksum=33F3F2F70774600332307DE5D206E96B [21 October, 2013].</p>		<p>education. It considers the effects of various aspects of qualification design and looks at how countries choose which options are most suitable to their own situation. Of particular interest to our report are the sections on centralised versus decentralised assessment systems, the role of standards in qualification frameworks, and grading issues at upper secondary level.</p>	<p>but the overlap between the countries considered in this paper and in our report is only partial.</p>
<p>Korobko, O.B., Glas, C.A.W., Bosker, R.J. and Luyten, J.W. (2008). 'Comparing the difficulty of examination subjects with item response theory', <i>Journal of Educational Measurement</i> 45, 2, 139–157.</p>	<p>The Netherlands</p>	<p>This paper develops a model (using IRT) for comparing the difficulty of examination subjects. The best-fit model takes into account differing subject difficulty, differing student ability, the fact that ability in one subject may not be the same as ability in another subject and the fact that students can choose their own subjects.</p>	<p>Theoretical research. Of relevance to the discussion about comparability of subjects.</p>
<p>Lamprianou, I. (2009). 'Comparability of examination standards between subjects: an international perspective', <i>Oxford Review of Education</i>, 35, 2, 205–</p>	<p>Australia (New South Wales and Tasmania) Fiji Singapore New Zealand</p>	<p>This paper presents a series of case studies that discuss various ways in which different jurisdictions have attempted to ensure comparability between</p>	<p>Review article. Of some use, but there is little overlap between the countries Lamprianou discusses and the ones in our report.</p>

Reference	Jurisdiction	Summary	Type and quality of evidence
226.	Scotland Greece	subjects.	
Philips, D. (2012). 'The impact of research on the evolution of secondary qualifications in New Zealand.' Paper presented at the British Educational Research Association Annual Conference, University of Manchester, 4-6 September.	New Zealand	A description of the original design of the National Certificate of Educational Achievement (NCEA) in new Zealand and how it has been modified into its current shape.	Case study of how one assessment system has evolved as a result of research and feedback.
Schildkamp , K., Rekers-Mombarg, L.T.M. and Harms, T.J. (2012). 'Student group differences in examination results and utilization for policy and school development', <i>School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice</i> , 23 , 2, 229–255.	The Netherlands	An investigation into the discrepancies between school-based assessment and national assessment of students in the Netherlands. The researchers found that for some groups of students the discrepancies were outside the limit tolerated by the Dutch Inspectorate. However the schools made little use of this data to improve planning and teaching	Research article. Of some limited relevance
van Rijn, P.W., Béguin, A.A. and Verstralen, H.H.F.M. (2012). 'Educational measurement issues and implications of high stakes decision making in final	The Netherlands	A theoretical investigation into the affects of changing the criteria used to make the pass/fail decision for school leaving examinations in the Netherlands.	Theoretical research. Of limited relevance to our report apart from the discussion of the methods used to establish the 'equating constant' used to scale

Reference

Jurisdiction

Summary

Type and quality of evidence

examinations in secondary education in the Netherlands', *Assessment in Education: Principles, Policy & Practice*, **19**, 1, 117–136.

examination scores.

Appendix C – Search strategy

Search strategies were developed for all databases by using keywords from the database thesauri, supplemented with free-text search terms (denoted by the abbreviation 'ft'). All searches were limited to material published between 2007 and 2013 and in the English language.

Australian Education Index (AEI)

Searched via ProQuest, 17.09.13

AEI is produced by the Australian Council for Educational Research. It is an index to materials at all levels of education and related fields. Source documents include journal articles, monographs, research reports, theses, conference papers, legislation, parliamentary debates and newspaper articles.

- #1 Standards
- #2 Academic standards
- #3 Educational standards (ft)
- #4 Establishment of standards (ft)
- #5 Maintenance of standards (ft)
- #6 Educational change
- #7 Educational reform (ft)
- #8 Qualification reform (ft)
- #9 Examinations (ft)
- #10 Qualifications
- #11 High stakes examinations (ft)
- #12 High stakes qualifications (ft)
- #13 School leaving examinations (ft)
- #14 School leaving qualifications (ft)
- #15 Matriculation (ft)
- #16 Tests
- #17 Achievement tests
- #18 Reference testing (ft)
- #19 Criterion referenced tests
- #20 Norm referenced tests
- #21 Mixed methods (ft)
- #22 Monitoring
- #23 Monitoring (assessment)
- #24 Reporting (student achievement)



- #25 Grading
- #26 Grading procedures (ft)
- #27 Grading system (ft)
- #28 Examination boards (ft)
- #29 Examination regulators (ft)
- #30 Advanced placement
- #31 Abitur (ft)
- #32 Alberta High School Diploma (ft)
- #33 American College Testing (ft)
- #34 Baccalauréat (ft)
- #35 Certificate of Graduation from Upper Secondary Education (ft)
- #36 College Scholastic Ability Tests (ft)
- #37 Diploma of High School Education (ft)
- #38 Hong Kong Diploma of Secondary Education (ft)
- #39 Massachusetts Comprehensive Assessment System (ft)
- #40 National Assessment of Academic Ability (ft)
- #41 National Certificate of Educational Achievement (ft)
- #42 NCEA (ft)
- #43 Pan Canadian Assessment Programme (ft)
- #44 Provincial Achievement Tests (ft)
- #45 Queensland Certificate of Education (ft)
- #46 QCE (ft)
- #47 SAT Reasoning Tests (ft)
- #48 Singapore-Cambridge General Certificate of Education (ft)
- #49 Studentexamen (ft)
- #50 Voorbereidend Wetenschappelijk Onderwijs (ft)
- #51 VWO (ft)
- #52 Ylioppilastutkinto (ft)
- #53 #1 OR #2 OR #3 ... OR #52

- #54 Alberta (ft)
- #55 Finland (ft)
- #56 France (ft)
- #57 Germany (ft)
- #58 Hong Kong (ft)
- #59 Japan (ft)
- #60 Massachusetts (ft)



- #61 Netherlands (ft)
- #62 New Zealand (ft)
- #63 Queensland (ft)
- #64 Singapore (ft)
- #65 South Korea (ft)
- #66 #54 OR #55 OR #56 ... OR #65

#67 #53 AND #66

British Education Index (BEI)

Searched via ProQuest, 17.09.13

BEI provides information on research, policy and practice in education and training in the UK. Sources include over 300 journals, mostly published in the UK, plus other material including reports, series and conference papers.

- #1 Standards
- #2 Academic standards
- #3 Educational standards (ft)
- #4 Establishment of standards (ft)
- #5 Maintenance of standards (ft)
- #6 Examinations
- #7 Qualifications
- #8 Educational change
- #9 Educational reform (ft)
- #10 Qualification reform (ft)
- #11 High stakes examinations (ft)
- #12 High stakes qualifications (ft)
- #13 School leaving examinations (ft)
- #14 School leaving qualifications (ft)
- #15 Matriculation (ft)
- #16 Achievement tests
- #17 Reference testing (ft)
- #18 Criterion referenced tests
- #19 Norm referenced tests
- #20 Mixed methods (ft)
- #21 Monitoring
- #22 Standards referenced reporting (ft)
- #23 Marking (scholastic)



- #24 Grading procedures (ft)
- #25 Grading system (ft)
- #26 Examination boards
- #27 Examination regulators (ft)
- #28 Advanced placement
- #29 Abitur (ft)
- #30 Alberta High School Diploma (ft)
- #31 American College Testing (ft)
- #32 Baccalauréat (ft)
- #33 Certificate of Graduation from Upper Secondary Education (ft)
- #34 College Scholastic Ability Tests (ft)
- #35 Diploma of High School Education (ft)
- #36 Hong Kong Diploma of Secondary Education (ft)
- #37 Massachusetts Comprehensive Assessment System (ft)
- #38 National Assessment of Academic Ability (ft)
- #39 National Certificate of Educational Achievement (ft)
- #40 NCEA (ft)
- #41 Pan Canadian Assessment Programme (ft)
- #42 Provincial Achievement Tests (ft)
- #43 Queensland Certificate of Education (ft)
- #44 QCE (ft)
- #45 SAT Reasoning Tests (ft)
- #46 Singapore-Cambridge General Certificate of Education (ft)
- #47 Studentexamen (ft)
- #48 Voorbereidend Wetenschappelijk Onderwijs (ft)
- #49 VWO (ft)
- #50 Ylioppilastutkinto (ft)
- #51 #1 OR #2 OR #3 ... OR #50

- #52 Alberta (ft)
- #53 Finland (ft)
- #54 France (ft)
- #55 Germany (ft)
- #56 Hong Kong (ft)
- #57 Japan (ft)
- #58 Massachusetts (ft)
- #59 Netherlands (ft)



- #60 New Zealand (ft)
- #61 Queensland (ft)
- #62 Singapore (ft)
- #63 South Korea (ft)
- #64 #52 OR #53 OR #54 ... OR #63

#65 #51 AND #64

Education-line

Searched 19.09.13

The free collections search interface of the British Education Index provides access to a unique collection of over 6,000 texts, mostly but not exclusively, conference papers, presented to the BEI by their authors. The collection also covers work done by the Teaching and Learning Research Programme.

- #1 Standards
- #2 Academic standards
- #3 Educational standards (ft)
- #4 Establishment of standards (ft)
- #5 Maintenance of standards (ft)
- #6 Examinations
- #7 Qualifications
- #8 Educational change
- #9 Educational reform (ft)
- #10 Qualification reform (ft)
- #11 High stakes examinations (ft)
- #12 High stakes qualifications (ft)
- #13 School leaving examinations (ft)
- #14 School leaving qualifications (ft)
- #15 Matriculation (ft)
- #16 Achievement tests
- #17 Reference testing (ft)
- #18 Criterion referenced tests
- #19 Norm referenced tests
- #20 Mixed methods (ft)
- #21 Monitoring
- #22 Standards referenced reporting (ft)
- #23 Marking (scholastic)



- #24 Grading procedures (ft)
- #25 Grading system (ft)
- #26 Examination boards (ft)
- #27 Examination regulators (ft)
- #28 Advanced placement
- #29 Abitur (ft)
- #30 Alberta High School Diploma (ft)
- #31 American College Testing (ft)
- #32 Baccalauréat (ft)
- #33 Certificate of Graduation from Upper Secondary Education (ft)
- #34 College Scholastic Ability Tests (ft)
- #35 Diploma of High School Education (ft)
- #36 Hong Kong Diploma of Secondary Education (ft)
- #37 Massachusetts Comprehensive Assessment System (ft)
- #38 National Assessment of Academic Ability (ft)
- #39 National Certificate of Educational Achievement (ft)
- #40 NCEA (ft)
- #41 Pan Canadian Assessment Programme (ft)
- #42 Provincial Achievement Tests (ft)
- #43 Queensland Certificate of Education (ft)
- #44 QCE (ft)
- #45 SAT Reasoning Tests (ft)
- #46 Singapore-Cambridge General Certificate of Education (ft)
- #47 Studentexamen (ft)
- #48 Voorbereidend Wetenschappelijk Onderwijs (ft)
- #49 VWO (ft)
- #50 Ylioppilastutkinto (ft)
- #51 #1 OR #2 OR #3 ... OR #50
- #52 Alberta (ft)
- #53 Finland (ft)
- #54 France (ft)
- #55 Germany (ft)
- #56 Hong Kong (ft)
- #57 Japan (ft)
- #58 Massachusetts (ft)
- #59 Netherlands (ft)
- #60 New Zealand (ft)



- #61 Queensland (ft)
- #62 Singapore (ft)
- #63 South Korea (ft)
- #64 #52 OR #53 OR #54 ... OR #63

- #65 #51 AND #64

Education Resources Information Center (ERIC)

Searched via ProQuest, 18.09.13

ERIC is sponsored by the United States Department of Education and is the largest education database in the world. Coverage includes research documents, journal articles, technical reports, programme descriptions and evaluations and curricula material.

- #1 Standards
- #2 Academic standards
- #3 Educational standards (ft)
- #4 Establishment of standards (ft)
- #5 Maintenance of standards (ft)
- #6 Examinations (ft)
- #7 Exit examinations
- #8 Qualifications
- #9 Educational reform (ft)
- #10 Qualification reform (ft)
- #11 High stakes tests
- #12 High stakes examinations (ft)
- #13 High stakes qualifications (ft)
- #14 School leaving examinations (ft)
- #15 School leaving qualifications (ft)
- #16 Matriculation (ft)
- #17 Admission (school)
- #18 Tests
- #19 Achievement tests
- #20 Reference testing (ft)
- #21 Criterion referenced tests
- #22 Norm referenced tests
- #23 Mixed methods (ft)
- #24 Grading
- #25 Grading procedures (ft)



- #26 Grading system (ft)
- #27 Standards referenced reporting (ft)
- #28 Progress monitoring
- #29 Examination boards (ft)
- #30 Examination regulators (ft)
- #31 Advanced placement
- #32 Abitur (ft)
- #33 Alberta High School Diploma (ft)
- #34 American College Testing (ft)
- #35 Baccalauréat (ft)
- #36 Certificate of Graduation from Upper Secondary Education (ft)
- #37 College Scholastic Ability Tests (ft)
- #38 Diploma of High School Education (ft)
- #39 Hong Kong Diploma of Secondary Education (ft)
- #40 Massachusetts Comprehensive Assessment System (ft)
- #41 National Assessment of Academic Ability (ft)
- #42 National Certificate of Educational Achievement (ft)
- #43 NCEA (ft)
- #44 Pan Canadian Assessment Programme (ft)
- #45 Provincial Achievement Tests (ft)
- #46 Queensland Certificate of Education (ft)
- #47 QCE (ft)
- #48 SAT Reasoning Tests (ft)
- #49 Singapore-Cambridge General Certificate of Education (ft)
- #50 Studentexamen (ft)
- #51 Voorbereidend Wetenschappelijk Onderwijs (ft)
- #52 VWO (ft)
- #53 Ylioppilastutkinto (ft)
- #54 #1 OR #2 OR #3 ... OR #53
- #55 Alberta (ft)
- #56 Finland (ft)
- #57 France (ft)
- #58 Germany (ft)
- #59 Hong Kong (ft)
- #60 Japan (ft)
- #61 Massachusetts (ft)
- #62 Netherlands (ft)



- #63 New Zealand (ft)
- #64 Queensland (ft)
- #65 Singapore (ft)
- #66 South Korea (ft)
- #67 #55 OR #56 OR #57 ... OR #66
- #68 #54 AND #67



NFER provides evidence for excellence through its independence and insights, the breadth of its work, its connections, and a focus on outcomes.

- independent
- insights
- breadth
- connections
- outcomes

© 2013 National Foundation for Educational Research

**National Foundation for
Educational Research**
The Mere, Upton Park
Slough, Berks SL1 2DQ

T: 01753 574123
F: 01753 691632
E: enquiries@nfer.ac.uk
www.nfer.ac.uk

OFQN