

Is it valid to treat assessment grades from different subjects the same?

Dennis Opposs, Standards Chair, *Office of Qualifications and Examinations Regulation*, (Ofqual), Coventry CV5 6UB, UK. dennis.oppo@ofqual.gov.uk

Abstract

Some say that it is meaningless to compare standards in different subjects to each other. But what if those standards are actually different but become a common currency so that, for example, a grade A in mathematics is treated in the same way as a grade A in geography? Is that interpretation and use of the grades valid?

Ofqual is trying to understand better how different jurisdictions deal with comparability between subjects in assessments used for university entrance. England is unusual in that pre-university students have an almost free choice of subjects and take few subjects in total – typically just three. In jurisdictions where the curriculum is constrained or where all those competing for the same university places take a very similar assessment, comparability between subjects may be of little interest. There are though some jurisdictions where there is more curriculum choice and where an important feature of their university entrance systems are statistical adjustments to subject outcomes even though these may be controversial.

This paper explores how different jurisdictions make or do not make adjustments to grades across subjects. The session at the conference is intended to provide an opportunity for those from different countries to share their knowledge and experiences in this area.

Keywords: assessment; subjects; scaling; standards comparability; university entrance

Introduction

Ofqual is investigating whether some GCE A level¹ subjects can be considered harder or easier than others. If they can, we want to decide whether it would be beneficial to produce a new alignment and how that might be done. This paper is based on what we have learned from practices in other jurisdictions that do or do not make adjustments aimed at aligning different subjects.

We are seeking and will very much welcome feedback from in-country experts, those much closer to the operation of individual assessments referenced in this paper. That should allow us in due course to publish a fuller paper in which we have more confidence.

Methodology

Assessments

This review focuses on inter-subject comparability in high stakes assessments - gateway assessments that enable students to access the next stage of education or employment.

In some jurisdictions, entrance into almost all higher education institutions (HEIs) at

¹ GCE A levels - General Certificate of Education Advanced levels, normally taken at the end of upper secondary education at age 18. Each A level is an assessment in a particular subject. Over 45 subjects are available. There is no compulsory subject element – A levels can be taken in any combination desired to reflect the interests (or intended progression) of the student. A typical A level student takes three or four subjects.

undergraduate level requires students to take one particular assessment. An example of this is the *Gāokǎo* (National Higher Education Entrance Examination) used in China.

In other jurisdictions there may be more choice. So, for example, in New Zealand, students in some schools are prepared for the International Baccalaureate, students in others are prepared for international A level examinations, while students from most schools in the country take the national examination – the National Certificate of Educational Achievement (NCEA). All of these assessments would then be used directly in the university entrance process.

Another way in which choice happens is exemplified by the United States of America (USA). Here each university judges high school students on the basis of its own criteria. These might include ACT scores or SAT scores or neither. Students wanting to enter a university might or might not take one of those assessments. However, some states, such as New York and Massachusetts, have their own state-wide examinations that, alongside Grade Point Averages and other information, are particularly critical for students wanting to enter university.

Some of the assessments covered are based closely on the curriculum that the students study in senior secondary school, for example, the Leaving Certificate Examinations in Ireland. Others are more reasoning or aptitude tests, for example, the Psychometric Entrance Test used in Israel for university entrance.

In some jurisdictions the assessment described here appears to be the sole criterion used for university entrance (for example, China). In some, there are in addition university based tests as well (for example, Japan). In others, national exams or school-based assessments also contribute (for example, Israel).

In this study, the rationale for selecting an initial sample of assessments focused on the jurisdictions within which they are taken. We were guided by the following criteria:

- Jurisdictions that were identified as high performing in international benchmarking studies – PISA (2012), TIMSS (2011) and PIRLS (2011)
- Jurisdictions that were known to undertake specific methods of addressing inter-subject comparability
- Jurisdictions that have the greatest similarities in terms of assessment structure to England's system

Other jurisdictions were then added to give greater geographical coverage and to ensure the inclusion of assessments that are better described as university entrance aptitude tests rather than achievement tests related to a taught curriculum.

Thirty assessments were reviewed in total.

Australia: New South Wales Higher School Certificate (HSC)

Australia: Tasmanian Certificate of Education (TCE)²

Brazil: High School National Exam (ENEM)

Canada: Alberta High School Diploma

² All the Australian states and territories use scaling procedures to convert their end of school assessment outcomes into a score used as the main criterion for entry into most undergraduate courses in the country. The two states chosen here are therefore illustrative of the more general approach in Australia to inter-subject comparability adjustments.

China: *Gāokǎo* (National Higher Education Entrance Examination)
 Cyprus: Pan Cyprian Exam
 Fiji: Fiji School Leaving Certificate
 Finland: *Ylioppilaskirjoitukset / Studentexamen* (Matriculation Examination)
 France: *Baccalauréat général*
 Germany: *Abitur*
 Ghana, Liberia, Nigeria, Sierra-Leone and The Gambia: West African Senior School Certificate Examination
 Greece: Pan-Hellenic Exam
 Hong Kong: Diploma of Secondary Education (HKDSE)
 International Baccalaureate Diploma (IB Diploma)
 Ireland: Leaving Certificate Examinations
 Israel: Psychometric Entrance Test
 Japan: National Centre Test
 Kazakhstan: Unified National Test
 Netherlands: *Voorbereidend wetenschappelijk onderwijs*, or *vwo*
 New Zealand: National Certificate of Educational Achievement (NCEA)
 Poland: *Matura* (High School Examination)
 Russia: Unified state examination (EGE)
 Singapore: Primary School Leaving Examination (PSLE)
 South Africa: National Senior Certificate
 Switzerland: Federal Maturity Certificate
 Taiwan: The Basic Competency Test (BCT)
 Thailand: General Aptitude Test (GAT) and Professional Aptitude Test (PAT)
 UK: Scotland Standard Grade, Intermediate 1 and 2, Higher and Advanced Higher
 USA: SAT I and SAT II
 USA: ACT

The assessments included in this study are all end of upper secondary assessments that enable students to access higher education or employment apart from two end of primary school assessments which facilitate access to selective secondary schools.

In reviewing the jurisdictions' approaches to inter-subject comparability, it is important to consider the context of the assessments. Integral to a jurisdiction's approach is the educational framework which determines the structure of the assessment. We have therefore categorised the structures into three groups:

- **Free choice:** Students can select subjects of their choice to study and in which to be assessed (within this option there may be one or more compulsory subjects, but elective subjects are chosen from a broad menu).
- **Restricted framework:** Students can select subjects from pre-defined limited subject groups in which to be assessed.
- **Uniform subjects:** Students are all assessed in the same subjects.

Approach

The review was undertaken as desk research, focusing on publicly available information. The majority of the information was drawn from official education ministry and assessment agency websites. Where information could not be found on these sites (particularly information regarding public perceptions), news websites were included in the review. Other sources referred to in this paper include published research reports and journal articles.

Caveats

There are limitations to this approach and the caveats below should be considered when reading this paper.

- The findings are not definitive and have not been validated by the jurisdictions involved.
- In the majority of cases the sources publish limited detail on the methodologies of the approach taken and no detail on the rationale of selecting the approach or the impact it has had.
- It is not always clear whether the detail available is current.

Assessments where there is evidence that inter-subject comparability is addressed

Methods to address inter-subject comparability are implemented in some of the jurisdictions reviewed. Statistical modelling techniques are commonly applied in these jurisdictions, taking into account the relative difficulty of subjects when assessing the results of each student. In most cases this is in order to support access to higher education, though in the Singapore and Taiwan examples it is to support access to selective options within secondary education. A summary of the findings is provided in Table 1 below.

Table 1: Overview of assessments that address inter-subject comparability

Jurisdiction	Assessment	No. subjects	Assessed Subjects	Description	Method
Cyprus	Pan Cyprian exam	4	Free choice	Z-score scaling applied to the results to rank students regardless of subject choice.	Z-score scaling
Fiji	Fiji School Leaving Certificate	4	Free choice	Average marks scaling applied to the results to rank all students regardless of subject choice.	Average marks scaling
Hong Kong	Hong Kong Diploma of Secondary Education (HKDSE)	4	Restricted framework	Standards-referenced reporting and a Group Ability Index (GAI) are conducted with subject choice within a restricted framework.	Group Ability Index/ equipercentile model/ Rasch model
New South Wales, Australia	Higher School Certificate (HSC)	Varies	Free choice	Standards referenced assessment which uses average marks scaling. The Australian Tertiary Admission Rank (ATAR) is scaled.	Average marks scaling
Scotland, UK	Standard Grade, Intermediate 1 and 2, Higher and Advanced Higher	Varies	Free choice	Kelly's method is applied to produce national ratings for each subject..	Kelly's method
Singapore	Primary School Leaving Examination (PSLE)	4	Uniform subjects	T-Score scaling is applied to rank students in order of attainment.	T-score
Taiwan	The Basic Competency Test (BCT)	5	Uniform subjects	Item response theory (IRT) is applied to rank students by attainment.	Item response theory
Tasmania, Australia	Tasmanian Certificate of Education (TCE)	4	Uniform subjects	Rasch Analysis is used to produce scaled scores using the relative difficulty of each subject.	Rasch model

From the 30 assessments reviewed in total, eight jurisdictions explicitly address inter-subject comparability. Other jurisdictions may address inter-subject comparability although that is not clear from the information reviewed.

The assessment structure varies for each of the eight jurisdictions identified.

- One jurisdiction limits subject choice within a restricted framework (Hong Kong).
- Three of the jurisdictions have uniform subject choices for assessments (Singapore, Taiwan, Tasmania) including the two jurisdictions where the assessment reviewed was to access secondary education (Singapore, Taiwan).
- Four allow free subject choice (Cyprus, Fiji, New South Wales, Scotland).
- Five of the jurisdictions examine four subjects (Cyprus, Fiji, Hong Kong, Singapore, Tasmania); one includes five subjects (Taiwan); and two of the jurisdictions vary the number of subjects based on student choices (New South Wales, Scotland).

The following statistical methods have been used to address inter-subject comparability in assessments identified in the review:

- Latent Trait Models:
 - Rasch model: Tasmania
 - IRT: Taiwan
- Common Examinee Linear Models:
 - Kelly's method: Scotland
 - Average Marks Scaling: New South Wales, Fiji
 - Scaling using T-Scores: Singapore
 - Scaling using Z-Scores: Cyprus
 - Equipercentile analysis: Hong Kong

The **Hong Kong** Diploma of Secondary Education (HKDSE) Examination uses multiple methods which are designed to address inter-subject comparability and maintain standards over time. Subjects are categorised into three groups; core, elective and applied subjects. In setting standards, judgemental methods, through the inspection of scripts and reference to level descriptors, and statistical methods are applied. Different statistical methods are applied to different subject groups to produce a set of recommended cut scores. To address inter-subject comparability in specific elective subjects and to assist in grading applied subjects, a group ability index (GAI) is calculated for each level based on the candidature's results in the four core subjects and the correlations between the subjects. Results are then adjusted accordingly. The four core subjects are also monitored annually with a representative group of selected schools. A latent trait model is applied to monitoring test data and live examination data to standardise all items in the different examinations and generate the suggested cut scores.

In **Tasmania**, the Rasch model is used to scale subject scores in the Tasmanian Certificate of Education (TCE) in order to generate a Tertiary Entrance Score for each student to enable them to access higher education. In order to make comparisons between subjects, Tasmanian authorities assume that all subjects are underpinned by a common construct of 'general academic ability' or 'merit to enter university'. Rasch analysis of whole subject assessments rather than items is undertaken for every subject, and each subject is equated onto a common scale at three award points (satisfactory achievement, high achievement and outstanding

achievement). The model takes into account all the subjects undertaken by students and the award threshold positions are adjusted on the scale according to the relative difficulty of the subjects. Once the analysis is complete and the scaled thresholds for each subject have been finalised, the scores in between the threshold positions are filled in and a combined score on the common scale is produced for each student. This ensures that the scaled subject scores are directly comparable.

New South Wales applies average marks scaling to Higher School Certificate results and makes adjustments to generate an Australian Tertiary Admission Rank (ATAR) score for each student to enable them to access higher education. English is the only compulsory subject and students can choose from over 100 courses to complete their HSC. The scaling approach is based on the principle that when a common candidature takes two or more of the same subjects, then the average performance of the group should be roughly the same. The results of each group of students (common candidature) in every possible pair of subjects are compared and the raw scores are then scaled so that the results in different subjects are adjusted to take into account the difficulty of the subject. This combined score forms the ATAR.

In **Scotland**, the Scottish Qualifications Authority (SQA) annually produces national ratings based on a similar approach to that used in New South Wales. They employ Kelly's Method to compare grades achieved by candidates in one subject with how they performed in all other subjects to estimate the difficulty of that subject - the national rating. These indicate how many grades higher or lower the candidate group achieved in a subject than they achieved on average in their other subjects. Although no longer published, the ratings are still considered during the development of assessments and are discussed at the meetings where grade thresholds are determined.

In **Cyprus**, a Z-score scaling method is applied to convert the raw scores of subjects in their pre-university examinations. A standard deviation of 3 and a mean of 10 are applied to rescale all the scores and an aggregate score is calculated for selection purposes.

Similarly, in **Fiji**, average marks scaling is used and standardised scores are reported. Raw scores are converted onto a scale with a mean of 50 and standard deviation of 17. The mean and standard deviation are set centrally by the Ministry of Education based on the performance of previous cohorts.

Singapore applies a scaling method to the Primary School Leaving Examination (PSLE) subjects which ranks all students according to their performance. Students' proficiency in English language, one language selected from a prescribed range of "mother tongue" languages (Chinese, Malay and Tamil), mathematics and science is nationally examined. Generally, students are able to take subjects at either foundation or higher level. In each mother tongue language there are three levels of examination; standard, foundational and 'Higher Mother Tongue' level. Due to the varying raw marks between assessments, scores are converted to T-scores where the mean is 50 and the standard deviation is 10 in each subject. An aggregate score is then produced to assist in secondary school selection.

The Basic Competency Test (BCT) at the end of primary school exam in **Taiwan** uses IRT models to convert raw scores to scaled scores. The multiple-choice tests taken over two days comprise six subjects: Chinese, English, mathematics, science, social science, and writing

and each has a scaled score ranging from 1-60 points. There are two opportunities each year to sit the test with students given a reported score out of 300 and a percentage ranking (1-99).

Public perceptions of addressing inter-subject comparability through statistical methods

It is important that the conduct of an examination system is perceived as fair and acceptable to the public in order to be trusted and promote confidence in the results. The use of statistical methods by some of the jurisdictions reviewed has led to critical comments from public, professional and academic sources. These highlight a risk that using statistical methods that are unintelligible to most audiences to align subjects can result in mistrust and a lack of confidence in the system.

Additionally, scaling scores can influence student choice. Students and teachers can try to devise methods to ‘work’ the system. For example, they might identify ‘easy’ subjects – those that are expected to be scaled down. They then devise specific (favourable) combinations of subjects in an attempt to avoid their results being scaled down.

Some examples of these issues are given below.

There is widespread concern in **Australia** about the reduction in the number of students studying calculus-based mathematics courses in the final year of secondary education. Recent research indicates that on average, those in New South Wales who study general mathematics as part of their Higher School Certificate (HSC) achieve materially higher scaled scores than those who undertake the calculus-based course. “The current scaling mechanism provides a strong incentive to take HSC general mathematics for a very large group of students. At a time when many are deeply concerned about the reducing numbers of students studying higher level mathematics in the final year of secondary education, it is useful to consider the evidence presented here which supports one of the possible explanations for this drop in numbers” (Pitt 2015 p80).

Of more than 1000 mathematics teachers surveyed, half believed that some students in their school were selecting senior mathematics courses below their capability. A desire to optimise HSC and ATAR results was the most common reason given for these selections and was cited over 200 times by these teachers. (Pitt 2015)

A newspaper article (Sydney Morning Herald, 2015) based on the research generated many on-line comments, some of which indicated that the implications of the scaling system go much wider than choices of mathematics courses and do relate to inter-subject comparability. For example, “I went to an HSC information night. . . Parents and students were very concerned about scaling both within and between subjects. Some students spoke of the difficulty of deciding . . . trying to take into account their own ability and the perceived scaling of that level of the subject. . . every part mark counts when you are trying to enter courses.”

In 2013 in his speech at the National Day Rally, the Prime Minister of **Singapore** spoke about the T-score system used in the PSLE.

“The PSLE, everybody thinks it matters, heaven and earth. I do not know what my PSLE grade is. . . But today, it is different. . . Not just everybody knows his T-score, everybody knows his friends’ T-score and his friends’ sons or daughters’ T-score. . . One-point difference in the PSLE scores, 230 versus 231, may make all the difference in your secondary

school posting. But at the age of 12, one examination, four papers and you want to measure the child to so many decimal points and say well, this one got one point better than that child? It is a distinction which is meaningless and too fine to make. Who is going to grow up abler, more committed, more capable, a better contributor to society? At the age of 12, you can guess, you cannot tell. Certainly, you cannot tell based on one point difference and I do not think we should decide secondary school postings based on such fine distinctions.

So we will score PSLE differently. We will use wider bands for grades, 'O' levels are like that . . . A1 to 9 . . . I think if we have a system of grades like that rather than precise scores, it will reduce the excessive competition to chase that last point. If you get an A* that is an A*, it does not matter where it is 91 A* or 99 A*. It is an A* and that is good enough." (Prime Minister's Office, 2013)

Although the issue raised in the speech is not specifically inter-subject comparability, the purpose of the scaling to produce the T-score is to align subject scores so that they can be aggregated.

In **Cyprus** since 2006, the raw results of the upper secondary school graduation exams (the Pancyprian Examinations) in 'easy' subjects have been scaled down and those in 'difficult' subjects scaled up to provide comparable access scores for university entrance purposes. One consequence is that students try to avoid subjects which historically are scaled down. For example, entries for chemistry dropped by 70% following the introduction of the scaling system (Lamprianou, 2007).

The media became interested because to the public, some of the stories about particular students appeared inexplicable. Some students with the same raw scores received very different scaled scores because they had taken different subjects. There were stories about students who took the same subjects, and had the same average raw score although different raw scores per subject. Depending on the statistical difficulty of each subject, some of these students then ended up with very different scaled scores.

Parents and students questioned the fairness of the system and even the accuracy of the calculations. Although the purpose of scaling is to adjust the raw scores because of the differential difficulty of the examination subjects, the end result puzzled parents, students and the press. (Lamprianou, 2012)

In **Fiji** there have been recurring reports of distrust in the scaling of marks. In 2006, students and parents complained to the Fiji Human Rights Commission that the scaling of exam marks was unfair and not transparent. In 2008, the Fijian Teachers Association requested a review of the scaling policy as they felt it was confusing for students and seemed to scale down able students while less able students were being scaled up.

In 2010, the Teachers Union stated that they did not support the scaling of exam marks and felt it was a government exercise rather than in the best interests of the students. However, the Fiji Principals Association were in favour of scaling believing that it put all students on a level field and enabled comparability. The Ministry of Education defended the system saying that it is based on sound educational assessment principles and is used internationally.

In the Fijian Parliament in February 2015, the education minister outlined his case for the removal of scaling, saying the practice has "caused substantial damage to the education

system and graduates in the market.” Raw mark evidence from 2009 to 2014, showed that mean marks in Year 12 and 13 steadily declined. “Madam speaker, there was no other alternative but to remove or discontinue the scaling of marks . . .” (Fiji Times, 2015)

Prior to the introduction of the National Certificate of Educational Achievement (NCEA) a decade ago the **New Zealand** system adjusted students’ results to attempt to improve inter-subject comparability using percentile analysis. Subjects’ standard scores were adjusted so that the performance of its group of candidates was comparable to that of the group in their other subjects. The inter-subject scaling of marks was a percentile analysis process based on the 95th, 90th, 75th, 50th, 25th, 10th and 5th percentiles and was applied to the national distribution of the marks for a subject based on candidates who had entered three or more subjects. However, this approach also meant that only a certain number of students could pass the examination and a fixed number of students would receive a fail grade. The system was felt to be unfair on students, as their success was relative to the performance of others, and it was feared that the focus on inter-subject comparability masked overall changes in student performance over time. (Jones, Phillips and van Krieken, 2005).

These concerns led to the introduction of the NCEA, which is standards-related and credit-based. It allows students the flexibility to choose the subjects they want to study to gain credits towards their final certificate. When the NCEA was first implemented, the proportion of results awarded at each achievement level (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. More than 10 years after the introduction, debate continues about whether some students are being deliberately directed into “soft” subjects. (New Zealand Herald, 2015)

Summary

The jurisdictions identified utilised a range of statistical approaches in attempting to address inter-subject comparability. The jurisdictions are diverse in the composition of their assessment systems, with variance in structure, number of subjects, exam approaches and marking.

The eight jurisdictions identified as implementing attempts to address inter-subject comparability all used statistical methods. This may be due to the necessity of publishing the calculation and approach to ensure transparency and confidence in the education system. Other jurisdictions may attempt to address inter-subject comparability, for example, during assessment design through the use of judgements, but we have not been able to find such information through publicly available websites.

Assessments with limited or no evidence that inter-subject comparability is addressed

There was little or no evidence that the other jurisdictions reviewed implement approaches to improve inter-subject comparability. However, due to the limitations of this review, it may be that the jurisdictions do attempt to improve inter-subject comparability but that the information is not publicly available. There is some evidence that jurisdictions that apply weightings to particular subjects are using judgements to decide these weightings. However, it is unclear from the evidence available whether these are due to subject difficulty, demand or a variety of other factors (such as teaching hours). In all cases, the assessments reviewed were used at the end of upper secondary school to gain access to university.

From the 30 assessments reviewed, 22 assessments do not appear to address inter-subject comparability based on the evidence available. The findings are summarised in Table 2 below.

Table 2: Overview of assessments that do not address inter-subject comparability

Jurisdiction	Assessment	No. subjects	Assessed Subjects	Description
Alberta, Canada	High School Diploma	Varies	Free choice	Students must achieve 100 credits made up of six mandatory and some elective subjects. Most courses are each five credits.
Brazil	High School National Exam (ENEM)	5	Uniform subjects	Increasingly used in Brazil to gain university entrance. Comprises 180 multiple-choice questions in five main areas: natural sciences, maths, human sciences, Portuguese and a foreign language. Students are also required to write an essay. The exam is scored out of 1,000 points.
China	<i>Gāokǎo</i> (National Higher Education Entrance Examination)	4	Restricted framework	In most provinces students take Chinese, mathematics and a foreign language (generally English) and either the humanities suite or the science suite. The mandatory and elective subjects are given different predefined points values. Total score out of 750.
Finland	Matriculation Examination	4	Restricted framework	Subject choice is limited within a framework. Each subject is graded from 1 to 7.
France	<i>Baccalauréat</i>	6	Restricted framework	Students select one of three series, within which subject choices are weighted differently depending on the series selected. Each subject is marked out of 20 with 10 being the minimum pass.
Germany	<i>Abitur</i>	10	Restricted framework	Subjects are divided into three areas, which are single, double or triple weighted in the final score. The <i>Abitur</i> uses a 15 point grading scale using numbers.
Ghana, Liberia, Nigeria, Sierra-Leone, The Gambia	West African Senior School Certificate Examination (WASSC)	8-9	Restricted framework	Multiple choice plus essays. Subject choice is limited within a framework. There is a nine point grading system from A1 (excellent) through C4–C6 (credit/minimum acceptable pass) to F9 (fail).
Greece	Pan-Hellenic exam	16	Restricted framework	Students select from one of four predefined pathways. Each subject is marked out of 20, 10 being a pass.
International Baccalaureate	Diploma programme	6	Restricted framework	Students select subjects from six subject groups which can be taken at standard or higher level. Subject grades range from 1 to 7. A student's final score is made up of the combined scores for each subject. The diploma is awarded to students who gain at least 24 points,
Ireland	Leaving Certificate Examinations	~7	Free choice	With the exception of Irish, students are able to choose which subjects they study although English and mathematics are effectively compulsory, and the majority of students take a third language. There are 13 grades from A1 to F.
Israel	Psychometric Entrance Test	3	Uniform subjects	The PET covers three areas: quantitative reasoning, verbal reasoning and the English language. One writing task plus 124 multiple-choice questions. The scoring scale ranges from 200 to 800 points.
Japan	National Centre Test	5	Uniform subjects	There are a total of 29 multiple-choice tests in six subjects. Candidates take the subjects specified by their university. Most subjects are each scored out of 100 points.

Kazakhstan	Unified National Test (UNT)	5	Restricted framework	125 multiple-choice questions. Exam covers five subjects: Kazakh language, Russian language, mathematics and Kazakh history plus an option – normally biology, physics or geography. The UNT is scored from 0 to 100; this is then converted to a grade of 2 to 5.
Netherlands	VWO	9	Restricted framework	Students select from one of four predefined pathways, in addition to mandatory general education subjects. Each subject is graded from 1 to 10 with an average final grade of 6 being the lowest pass.
New Zealand	National Certificate of Educational Achievement (NCEA)	Varies	Free choice	When students achieve the standards in a subject, they gain credits and once they have enough credits they get a NCEA certificate. A single achievement standard generally attracts three to four credits and a single subject usually has five to eight such standards; credits needed per course of study are usually 18 to 25. Students do not get an overall grade for a subject.
Poland	High school examination (Matura)	3	Restricted framework	Subject choice within a restricted framework. Percentage and percentile results are reported to compare results on a national scale.
Russia	Unified state examination (EGE)	4+	Restricted framework	Russian and mathematics compulsory. Optional tests in foreign languages, physics, chemistry, biology, geography, literature, history, social sciences and computing science. Multiple choice plus written answers required.
South Africa	National Senior Certificate	7+	Restricted framework	Seven subjects, including two compulsory official South African languages, either Mathematics or Mathematical Literacy, Life Orientation and three elective subjects. Grading of subjects is on a seven point rating scale, where 4 is the minimum acceptable pass.
Thailand	General Aptitude Test (GAT) and Professional Aptitude Test (PAT)		Restricted framework	The compulsory GAT covers reading, writing, analytical thinking, problem solving and English communication. The PAT has a choice of seven subjects – maths, science, engineering, architecture, education, arts and languages. The GAT and each PAT is scored out of 300 points.
Switzerland	Federal Maturity Certificate	9	Restricted framework	Students all study mandatory subjects, with an elective subject of focus and a supplementary subject. Subjects are weighted according to teaching hours. Each subject is graded, 6 being the maximum grade.
USA	SAT I	Varies	Uniform subjects	There are two versions of the SAT test. SAT I tests measure general verbal and quantitative reasoning. They comprise three sections: writing, reading and mathematics, the last two of which are primarily multiple choice. Possible scores range from 600 to 2400.
USA	SAT II	Varies	Free choice	SAT II tests – which far fewer students take – are subject-based. Students typically take three subjects chosen from the 20 available. Each test is scored from 200 to 800.
USA	ACT	Varies	Uniform subjects	The subject-based ACT consists of four multiple-choice tests: English, mathematics, reading and science (each scored from 1 to 36), with an optional writing section (scored from 1 to 12).

The assessment structures vary.

- Most limit subject choice within a restricted framework (examples include the International Baccalaureate and the West African SSC)
- Others have uniform subject choices (including Japan and Brazil)
- A minority allow free subject choice (including Alberta and New Zealand).

The following assessment system structures were evident.

- Subject choice limited within a restricted framework:
 - Units/subjects arranged in prescribed subject groups from which students must select a subject per group (for example, Finland and Switzerland).
 - Pathway approach that pre-defines subject combinations (for example, Greece and The Netherlands).
 - Application of pre-defined weighting to different subject areas (for example, France and Germany).
- Uniform subject choices for examinations:
 - Set national/state exam that all students take (for example, Japan).
- Free subject choice:
 - Credit based system that assigns a prescribed number of credits to each subject (Alberta and New Zealand).

Jurisdictions where no evidence could be found as to whether inter-subject comparability approaches were implemented do not explicitly state why they do not address the issue. In this regard, we could assume, for example, that where subject choice is offered within a restricted framework, subjects are categorised and selection is controlled so that the design of the education system means that students take similar combinations of subjects, making attainment broadly comparable - but it is just an assumption.

In education systems with restricted frameworks there is the necessity for students to identify their preferred university course and choose the appropriate pathway early on. This ensures that all students applying for specific university courses will have a very similar assessment profile and will therefore be comparable within their field. Similarly, those jurisdictions that allow students to select from groups of subjects assume equivalency within or between groups and therefore there is an overall balance in the assessment profile of the students. In most cases there is the opportunity for students to take subjects at different levels or at different weightings within the restricted framework to differentiate between candidates. The selection of subjects from a restricted framework can ensure breadth through the necessity of studying subjects from disparate areas and depth by focusing/weighting particular subjects. It is again necessary for students to be aware of the subjects they need to study, and at which level/weighting, for their desired university course.

Education systems that use a credit based system provide the flexibility to gain more credits by selecting subjects that are perceived to be more challenging. So do those that apply weightings to subjects to inform the final grade. The process of defining the subject credits or weightings varies by country.

In **New Zealand**, the credits available per subject for the NCEA are based on curriculum standards which have been defined by subject experts.

In **France**, the subject weightings for the *Baccalauréat* depend on the importance of the subject to the pathway, the depth of the syllabus and the teaching hours. Science subjects receive a higher weighting if you select the scientific pathway and a lower weighting if you

select the literary pathway, and vice versa for literary subjects. For example, philosophy is a key subject for the literary pathway and it therefore has a wide-ranging syllabus, receives eight teaching hours per week and has a high weighting. In the scientific pathway, philosophy is not a key subject and is subsequently taught for two or three hours per week, covers a limited syllabus and receives a low weighting. This allows students to select the pathway that best suits their strengths and interests.

The **International Baccalaureate** allows students to study subjects at either higher or standard level, with the higher level options having a more considerable syllabus and increased teaching hours. In **Switzerland**, weightings are related to teaching hours. In **Germany**, the weighting is dependent on the point where assessment is taken with core subjects double weighted throughout the duration, other subjects single weighted and final examination subjects triple weighted.

One of the outcomes of inter-subject comparability approaches in the jurisdictions reviewed was to enable students to be placed on a common scale for selection purposes. This was also found to be evident in those jurisdictions that do not appear to apply methods of inter-subject comparability. **Finland** and **Poland** use norm-referencing to assign grades in school leaving examinations. Norm-referenced methods aid stakeholders in selecting the highest attaining students for higher education and employment opportunities.

Public perceptions of not addressing inter-subject comparability

Very little information could be found regarding public perceptions of inter-subject comparability in jurisdictions where there was no evidence of a statistical inter-subject comparability method being implemented. There was some evidence of debate around subject difficulty and which subjects were perceived to be ‘easier’ but this was limited. Concern was evident more generally around the format or management of exams rather than the comparability of subjects and some jurisdictions are undergoing reform of their assessments as a result of these concerns. Two examples are worthy of mention though.

The Leaving Certificate examination marks the end of upper secondary education in Ireland. It is taken by more than 90% of the age cohort. Although designed as a terminal examination for certification, in practice this purpose is overshadowed by its central role in selection decisions for HEIs. In the case of the great majority of applicants to most HEIs courses in Ireland, it is the sole criterion used in the selection decision. A discussion paper (Hyland 2011) listed 10 key concerns raised by various stakeholders, within and outside the education system. These concerns included:

- It is easier to get a high grade in some Leaving Certificate subjects than in others. Some students choose subjects because it is easier to get a high grade in them, rather than because of their relevance for the third level course for which they are applying.
- Some students choose their courses on the basis of their likely points rather than on their interest in the course – they don’t want to “waste their points”.

Reforms to the Leaving Certificate recently announced³ do not include any proposals to align subjects.

³http://www.transition.ie/files/Supporting%20a%20Better%20Transition%20from%20Second%20Level%20to%20Higher%20Education%20-%20Implementation%20and%20Next%20Steps_April%202015.pdf

In **China's** *Gāokǎo* a maximum of 750 points are available from the examinations. Chinese, mathematics and a foreign language are worth up to 150 points each, and there are a further 100 points for each subject (up to three subjects) in the humanities and science combinations. The overall mark received by the student is generally a weighted sum of their subject marks. The marks in the separate subjects are raw marks.

The government has announced that changes will be implemented in 2016. The weighting of English is to reduce from 150 to 100 points, and the weight of Chinese is being increased from 150 to 180 points. The reason for these changes has not been explicitly stated by the Ministry of Education but commentators suggest it is likely to be for two reasons. First, to reduce the disadvantage faced by candidates from low income backgrounds or rural settings who are likely to have less access to English compared to their higher income, city based peers. Second, the reduced weighting of English could be to favour other subjects like maths (150 points). In both cases this alteration in weighting is due to factors of subject equality and importance rather than subject difficulty or demand (Sinograduate, 2014).

Summary

Twenty-two assessments were reviewed where no evidence could be found as to whether they implement an approach to address inter-subject comparability. It may be that inter-subject comparability is addressed but that the information is not publicly available.

Conclusions

The purpose of the assessments reviewed was primarily to enable students to access higher education. In all jurisdictions, it was important for students to select the subjects necessary to access the university course they wished to study. The same is true of England.

Systems that use inter-subject comparability methods generally award 'certificate' style products where it is necessary to have both individual subject awards as well as an overall award. However, this is not exclusive to jurisdictions using inter-subject comparability statistical adjustments. Many that do not use inter-subject comparability methods also award certificate style products.

Those that use inter-subject comparability methods are also more likely to operate a free choice structure whilst those that appear not to use inter-subject comparability methods are more likely to operate a restricted framework approach although both structures are evident in each category. Systems that use statistical methods to address inter-subject comparability most commonly include four subjects whilst those that appear not to use inter-subject comparability methods most commonly include a higher number of subjects, ranging up to 16.

Statistical modelling of results to address inter-subject comparability can lead to distrust and reduced confidence in the system due to the complexity of the calculations applied, making them unintelligible to most audiences. This is evident from some public perceptions of the systems in Australia, Cyprus and Fiji in particular. It is not clear though whether confidence is necessarily higher in systems where there is some choice of subjects but no use is made of such statistical adjustments.

References

Fiji Times (2015) Available at: <http://www.fijitimes.com/story.aspx?id=295649>

Hyland, A (2011) *Entry to Higher Education in Ireland in the 21st Century*. Available at: [http://www.transition.ie/files/Entry to Higher Education in Ireland in the 21st Century% 20.pdf](http://www.transition.ie/files/Entry%20to%20Higher%20Education%20in%20Ireland%20in%20the%2021st%20Century%20.pdf)

Jones, B.E., Philips, D. and van Krieken, R. (2011) *Inter-Subjects Standards: An Insoluble Problem?* Manchester: Assessment and Qualifications Alliance.

Lamprianou, I. (2007) *Comparability methods and public distrust: An international perspective*. In *Techniques for monitoring the comparability of examination standards*, edited by: Newton, P., Baird, J., Goldstein, H., Patrick, H. and Tymms, p368–371. London: Qualifications and Curriculum Authority.

Lamprianou (2012) *Unintended consequences of forced policy-making in high stakes examinations: the case of the Republic of Cyprus*. *Assessment in Education: Principles, Policy & Practice* Volume 19, Issue 1, 2012 Special Issue: High-stakes testing - value, fairness and consequences

New Zealand Herald (2015) Available at: http://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=11430932

Pitt, D.G.W. (2015) *On the scaling of NSW HSC marks in mathematics and encouraging higher participation in calculus based courses*. *Australian Journal of Education* 2015, Vol. 59(1) 65–81

Prime Minister's Office (2013) Available at: <http://www.pmo.gov.sg/mediacentre/prime-minister-lee-hsien-loong-national-day-rally-2013-speech-english>

Sinograduate (2014) Available at: <http://www.sinograduate.com/comment/articles/why-china-reducing-importance-english-among-gaokao-exams>

Sydney Morning Herald (2015) Available at: <http://www.smh.com.au/national/education/hsc-maths-students-studying-advanced-maths-stung-with-lower-marks-in-atar-20150519-gh45ox.html>

Providers of the assessment systems selected

Alberta, Canada: Alberta Education, <http://www.education.alberta.ca>

Brazil: National Institute of Educational Studies and Research, portal.inep.gov.br/web/enem

China: Ministry of Education, www.moe.edu.cn

Cyprus: Ministry of Education and Culture, www.moec.gov.cy/ypexams/en/

Fiji: Ministry of Education www.education.gov.fj

Finland: Finnish National Board of Education, http://www.oph.fi/english/education/overview_of_the_education_system

France: Ministry of National Education, www.education.gouv.fr

Germany: Federal Ministry of Education and Research, www.bmbf.de/en/

Ghana, Liberia, Nigeria, Sierra-Leone and The Gambia: West African Examinations Council, <http://www.waecnigeria.org/Home.aspx>

Greece: Ministry of Education and Religious Affairs, www.minedu.gov.gr

Hong Kong: Hong Kong Examinations and Assessment Authority
[http://www.hkeaa.edu.hk/DocLibrary/HKCEE/Grading and Marking SRR/booklet_srr.pdf](http://www.hkeaa.edu.hk/DocLibrary/HKCEE/Grading_and_Marking_SRR/booklet_srr.pdf)

International Baccalaureate Organisation: www.ibo.org

Ireland: Department of Education and Skills, <https://www.education.ie/en/>

Israel: National Institute for Testing and Evaluation, <https://nite.org.il/index.php/en/>

Japan: Ministry of Education, Culture, Sports, Science and Technology (MEXT),
www.mext.go.jp/english/

Kazakhstan: National Testing Centre, <http://testcenter.kz/en/entrants/ent/>

The Netherlands: Ministry of Education, Culture and Science, www.government.nl
National Institute for Curriculum Guidance www.slo.nl

New South Wales, Australia:
[http://studentonline.bos.nsw.edu.au/go/seniorstudy/hsc rules and procedures/](http://studentonline.bos.nsw.edu.au/go/seniorstudy/hsc_rules_and_procedures/)

New Zealand: Ministry of Education, www.minedu.govt.nz ;
New Zealand Qualification Authority, www.nzqa.govt.nz

Poland: Central Examination Board, [http://apl-bud.home.pl/pdfs/edusystem and ext asesment Poland.pdf](http://apl-bud.home.pl/pdfs/edusystem_and_ext_asesment_Poland.pdf)

Russia: Federal Service for Supervision in Education and Science,
<http://government.ru/en/department/35/>

Scotland, UK: Scottish Qualifications Authority, <http://www.sqa.org.uk>

Singapore: MOE Singapore, <http://www.moe.gov.sg>

South Africa: Umalusi, <http://www.umalusi.org.za/>

Switzerland: Swiss Conference of Canton Ministers of Education (EDK),
www.edk.ch/dyn/11553.php

Taiwan: <http://www.wes.org/ewenr/10may/feature.htm>

Tasmania, Australia: Tasmanian Qualifications Authority, <http://www.tqa.tas.gov.au/1906>

Thailand: The National Institute of Educational Testing Service,
<http://www.niets.or.th/upload-files/uploadfile/5/5113f2fc40d9b7ccbf26972226c1a536.pdf>

USA: College Board, <https://sat.collegeboard.org/home> ; ACT,
<http://www.act.org/products/k-12-act-test/>