



**Cambridge
Assessment**

High-stakes testing after basic secondary education: How and why is it done in high-performing education systems?

Research Report

Irenka Suto

Tim Oates

February 2021

Author contact details:

Irenka Suto & Tim Oates

Assessment Research and Development
Research Division
Cambridge Assessment
The Triangle Building
Shaftesbury Road
Cambridge
CB2 8EA
UK

Suto.I@cambridgeassessment.org.uk

Oates.T@cambridgeassessment.org.uk

<http://www.cambridgeassessment.org.uk>

As a department of Cambridge University, Cambridge Assessment is respected and trusted worldwide, managing three world-class examination boards, and maintaining the highest standards in educational assessment and learning. We are a not-for-profit organisation.

Cambridge Assessment is committed to making our documents accessible in accordance with the WCAG 2.1 Standard. We're always looking to improve the accessibility of our documents. If you find any problems or you think we're not meeting accessibility requirements, contact our team: [Research Division](#)

If you need this document in a different format [contact us](#) telling us your name, email address and requirements and we will respond within 15 working days.

How to cite this publication:

Suto, I., & Oates, T. (2021). *High-stakes testing after basic secondary education: How and why is it done in high-performing education systems?* Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment.

Acknowledgements

We would like to thank Gill Elliott for cross-checking some of the data collated in this report. We are also grateful to Gill and to Nicky Rushton for their advice on method. Gillian Cooke kindly sourced material on the history of external assessment in England. Jing Zhao and Alistair Starling kindly provided information on assessment and accountability in China and Estonia respectively. This support has been invaluable.

Summary

In this report on education systems in Repeatedly High Performing Jurisdictions (RHPJs) we present data on the assessment approaches used at the end of basic secondary education. These assessments are conducted at around the age of 16, at approximately the stage when students in England, Wales and Northern Ireland take GCSEs.

Key findings

- No single approach to assessment at the end of basic secondary education is associated with the success of all RHPJs.
- Contrary to perceptions expressed in parts of the media, around two-thirds of RHPJs, including several from across Europe, utilise external assessment at the end of basic secondary education.
- In many of these jurisdictions external assessment plays critical roles in determining students' directions in upper secondary education, and in providing students with qualifications in subjects that they might never again study.
- This approach is rooted soundly in meritocratic principles.
- There is no evidence to suggest that abandoning external assessment, or not adopting it in the first place, is associated with higher student performances at a system level.
- Using assessment at the end of basic secondary education for the secondary purpose of accountability is not necessarily a bad thing, and England is not unique in this respect. Repeat high performers such as Estonia and Shanghai also do so.
- Both internal and external assessments at the end of basic secondary education can be 'high stakes' for students and for teachers too.

Background

Domestic discussions of reform frequently seek justification in evidence about high-performing systems around the world. Where such evidence is well grounded in reality, this can be of genuine value. However, when it is false, it can stimulate flawed shifts in both public sentiment and material arrangements. Regrettably, recent domestic discussion of examinations and assessment at the end of basic secondary education seriously misrepresents the situation globally. This misrepresentation is so extreme that significant problems may arise in domestic debate and policy formation.

Currently, different interests are converging on a 'Remove GCSEs' message (Baker, 2019; Lough, 2020; Rethinking Assessment, 2020; Richmond, 2021). Some are interested in dismantling accountability arrangements, some in pupil transfer at 14, some in student welfare, some in technological innovation, others in opening and liberalising a market in educational

assessment. Worryingly, some have publicised the misinformation that no other European countries have high-stakes testing at this age (Leaton Gray, 2018; Poole, 2020).

Context is important: arrangements in England were last reformed in 2015, focusing particularly on elevating standards in the GCSE (General Certificate of Secondary Education) and A Level (General Certificate of Education). It is extremely important to recognise that England's performance in PISA, TIMSS and PIRLS has increased since 2010. Recent governments have recognised the underdevelopment of a high-status vocational route, and acknowledge the importance of high-quality vocational education and training. Development of apprenticeship and technical qualifications is underway, intended to break the pattern of stalled initiatives in this area.

The long view also is important: GCSEs and A Levels have undergone significant innovation since their introduction (in 1986 and 1952 respectively), seeing innovation in assessment methods, grading structures, subject focus, and so on. This continues, with increased use of technology in administering and marking the examinations, enhanced assessment and direct support for learning.

This report is of course being released in the context of global school interruptions and closures due to the COVID-19 pandemic. Challenges in the delivery of summative assessment and certification for general and vocational qualifications in 2020 and 2021 have precipitated considerable domestic and international reflection and comment on the resilience of assessment and qualification arrangements. They have also accelerated discussion of digital enhancement and refinement of those arrangements - and of the linkages between assessment and learning. This report is intended to contribute materially to the evidence base for, and 'direction of travel' of, these discussions.

This report

The aim of this report is to bring some clarity and lesser-known facts to current debate. To this end, we present data collated in a desk-based study of all jurisdictions globally that have repeatedly performed highly in international comparisons. We reveal the forms of assessment that are used at the end of basic secondary education within each jurisdiction. This enables us to clarify how common high-stakes testing actually is within what are arguably the world's most successful education systems. Across the jurisdictions we then consider the stated purposes of assessments at this stage of education. This includes an exploration of earliest aspirations behind external examinations in England. In particular we examine the use of qualifications in progression, as this is of particular importance in current discussion. Since the use of assessment for accountability is perhaps its most controversial purpose, we explore this phenomenon in some renowned education systems: those of Shanghai in China, Estonia and Finland. Overall, we provide an evidence-based argument that when it comes to high-stakes testing at age 16, England, Wales and Northern Ireland are by no means as atypical as some would have us believe. Our approach to assessment is, in fact, strikingly similar to the approaches of many of the countries whose educational achievements are most admired.

This report builds on the international review, *Are claims that the GCSE is a white elephant red herrings?* (Elliott, Rushton, Darlington & Child, 2015), commissioned in 2014 by Tim Oates. The current report updates and extends the findings of the 'red herrings' report. Following a period of five years of educational developments around the world, it confirms the trends first identified in 2015.

In the identification of current RHPJs, we do not exhaustively examine the trajectory of performance of each system, for example, the decline of performance of Finland since 2006, or

the improvement in England in recent PISA and TIMSS cycles. For that we refer readers particularly to Crato (2020). Likewise, we do not exhaustively explore all the functions of the assessments and their accompanying materials and processes, such as conditioning and supporting didactics and pedagogy, specifying the content of learning programmes, lending structure to learning programmes, providing explicit and implicit quality assurance, motivating young people, and so on. For a full overview of the purposes of assessment we refer readers to Newton (2017), Coe & Heller-Sahlgren (2014) and Coles & Oates (2009).

Identification of Repeatedly High Performing Jurisdictions (RHPJs)

The most recent major international comparisons of students' educational performances are:

- Trends in International Mathematics and Science Study (TIMSS, 2019)
- Programme for International Student Assessment: Reading, Mathematics and Science (PISA 2018)
- Progress in International Reading Literacy Study (PIRLS, 2016)
- Pearson Index of Cognitive Skills and Educational Attainment (Pearson, 2014a, b).¹

In these international comparisons, the performances of all participating countries/jurisdictions are ranked. Adopting the approach taken by Elliott (2016) we began this study by identifying those jurisdictions that are *repeat* high performers according to these rankings. To do this, we identified the top 10 jurisdictions in each of the seven comparisons (Table 1).

¹ This index is a composite or 'basket' comparison, drawing partly from PISA 2012, TIMSS 2011 and PIRLS 2011 scores, and partly from literacy and graduation rates.

Table 1. Jurisdictions ranking in the top 10s of the seven most recent international comparisons

Rank	TIMSS 2019 (8th Grade Maths)	TIMSS 2019 (8th Grade Science)	PISA 2018 ^A (Reading)	PISA 2018 ^A (Maths)	PISA 2018 ^A (Science)	PIRLS 2016 (4th Grade Reading)	Pearson Index of Cognitive Skills and Educational Attainment 2014 ^B
1	Singapore	Singapore	B-S-J-Z ^C (China)	B-S-J-Z ^C (China)	B-S-J-Z ^C (China)	Russian Federation	South Korea
2	Chinese Taipei	Chinese Taipei	Singapore	Singapore	Singapore	Singapore	Japan
3	South Korea	Japan	Macao (China)	Macao (China)	Macao (China)	Hong Kong (China)	Singapore
4	Japan	South Korea	Hong Kong (China)	Hong Kong (China)	Estonia	Republic of Ireland	Hong Kong (China)
5	Hong Kong (China)	Russian Federation	Estonia	Chinese Taipei	Japan	Finland	Finland
6	Russian Federation	Finland	Canada	Japan	Finland	Poland	United Kingdom ^D
7	Republic of Ireland	Lithuania	Finland	South Korea	South Korea	Northern Ireland ^D	Canada
8	Lithuania	Hungary	Republic of Ireland	Estonia	Canada	Norway	Netherlands
9	Israel	Australia	South Korea	Netherlands	Hong Kong (China)	Chinese Taipei	Republic of Ireland
10	Australia	Republic of Ireland	Poland	Poland	Chinese Taipei	England ^D	Poland
Ranks in all seven top 10s: Singapore							
Ranks in six of the top 10s: Hong Kong (China), South Korea							
Ranks in five of the top 10s: Finland, Chinese Taipei (Taiwan), Republic of Ireland, Japan							
Ranks in four of the top 10s: Poland							
Ranks in three of the top 10s: Canada, Macao (China), Russian Federation, B-S-J-Z (China) ^C , Estonia, Belgium							
Ranks in two of the top 10s: England ^D , Northern Ireland ^D , Australia, Netherlands, Lithuania							
Ranks in one of the top 10s: Norway, Israel, Hungary							

Notes:

A: PISA students are aged between 15 years 3 months and 16 years 2 months at the time of the assessment, and they have completed at least six years of formal schooling.

B: The Pearson Index is a composite or 'basket' comparison which draws partly from the PISA 2012, TIMSS 2011 and PIRLS 2011 scores, and partly from literacy and graduation rates.

C: Beijing, Shanghai, Jiangsu and Zhejiang.

D: The UK's colour-coding reflects its inclusion of England and Northern Ireland. England appears in two top 10s: as England in one, and as part of the UK in one. Northern Ireland appears in two of the top 10s: as Northern Ireland in one, and as part of the UK in one. Scotland and Wales do not appear separately to the UK in any of the top 10s.

Table 1 shows that in total, 22 jurisdictions appear in at least one top 10.² Nineteen of these jurisdictions appear in two or more top 10s; we therefore termed them *Repeatedly High*

² We also analysed the top 20 jurisdictions in each international comparison. A table of top 20s equivalent to Table 1 includes 36 different countries/jurisdictions, 14 of which never rank at or above 10th place. This

Performing Jurisdictions (RHPJs). The table features many East Asian countries, with Singapore ranking within the top 10s (and even the top threes) of all seven international comparisons. Hong Kong and South Korea also do very well, ranking within six out of seven top 10s. The most successful European countries by this measure are Finland and the Republic of Ireland (five top 10s) followed by Poland (four top 10s), Estonia and Belgium (three top 10s). It is worth noting that RHPJs vary considerably in population size. Whilst South Korea has a similar population to that of England, Japan's and Russia's populations are much larger, as is the combined population of the four participating Chinese cities (232 million; see Schleicher, 2018). Many of the other jurisdictions are much smaller.

Collation of data on Repeatedly High Performing Jurisdictions

Next, we explored the assessment approaches used at the end of basic secondary education³ (at around age 16) in the 19 RHPJs. Adopting an established methodology for comparability research (Elliott, Rushton & Ireland, 2017), we collated relevant details of the 19⁴ jurisdictions' education systems from multiple reputable sources. We utilised major international comparative websites, including those of the OECD (2021), the European Union's Eurydice programme (European Commission, 2021), and UCAS's Qualification Information Profiles (UCAS, 2021). We also used the official websites of government ministries and associated assessment and education organisations. Wherever possible, we triangulated key information across sources.

We collated the data in good faith, but it should be noted that its pedigree can sometimes be difficult to ascertain. As Rushton and Elliott (2018) explain, although reputable organisations manage the major comparative databases, little is known about the various individuals who upload data into them. In the Eurydice programme, for example, each jurisdiction completes its own entry. Whilst this ensures a degree of authenticity, translation issues can sometimes blur the detail of the explanations, and some education systems do not fit neatly into the pre-set categories of description provided by the website hosts. Potentially, these issues can result in some ambiguity in interpretation. We therefore invited a highly experienced researcher to cross-check our data against the sources used, highlighting to her all areas of particular concern that required extra scrutiny. Through this process she identified no errors and very few ambiguities.

Assessment approaches in Repeatedly High Performing Jurisdictions

Once the data had been collated and checked, we used it to distinguish those RHPJs that use external assessment (national or regional) at the end of basic secondary education from those that do not. Table 2 shows that approximately two-thirds of the RHPJs use external assessment at this educational stage, either exclusively or in addition to internal assessment. Clearly, GCSEs are not an anomaly in this respect. In addition to England and Northern Ireland, the list of external assessment users includes four other European jurisdictions. These are the Republic of Ireland, Poland, Estonia and Belgium's French-speaking community. The number of European RHPJs that do *not* use external assessment is very similar. Overall, there is no striking geographical divide between the RHPJs that use external assessment and those that do not.

second 'tier' of countries/jurisdictions are: Bulgaria, Cyprus, Czech Republic, Denmark, Germany, Italy, Latvia, New Zealand, Portugal, Slovenia, Sweden, Switzerland, Turkey and the United States.

³ The term 'basic secondary education' is used throughout this report. Some jurisdictions use the terms 'lower' and 'main' secondary education instead. Very broadly (but with a few exceptions including England), this stage of education corresponds to ISCED Level 2 (UNESCO, 2011).

⁴ Although Belgium appeared as a single jurisdiction in the international comparisons, each of its three communities (French-speaking, German-speaking and Flemish-speaking) has a slightly different education system. We collated data on each of these separately.

Table 2. Repeatedly High Performing Jurisdictions with and without external assessment (national or regional) at the end of basic secondary education

Use external assessment (either exclusively or in addition to internal assessment)		Do not use external assessment	
Jurisdiction	Top 10 rankings in international comparisons	Jurisdiction	Top 10 rankings in international comparisons
Singapore	7/7	Hong Kong	6/7
South Korea	6/7	Finland	5/7
Republic of Ireland	5/7	Belgium ^A : German-speaking	3/7
Japan	5/7	Belgium ^A : Flemish-speaking	3/7
Chinese Taipei (Taiwan)	5/7	Macao (China)	3/7
Poland	4/7	Lithuania	2/7
Estonia	3/7	Netherlands	2/7
Belgium ^A : French-speaking	3/7		
Canada ^B : Ontario	3/7		
B-S-J-Z ^C (China)	3/7		
Russian Federation	3/7		
Australia ^D : NSW	2/7		
England	2/7		
Northern Ireland	2/7		

Notes:

A: Belgium comprises a French-speaking community, a German-speaking community and a Flemish-speaking community. The education systems of the three communities are similar in some respects but different in others.

B: Canada comprises 10 provinces and 3 territories, each with their own education systems. As it was beyond the scope of this study to analyse all the systems, that of the most populous province, Ontario, was selected.

C: Beijing, Shanghai, Jiangsu and Zhejiang.

D: Australia comprises 6 states and 10 territories, each with their own education systems. As it was beyond the scope of this study to analyse all the systems, that of the most populous state, New South Wales (NSW), was selected.

Purposes of assessments at the end of basic secondary education

Table 3 relates specifically to the 14 RHPJs that use external assessment after basic secondary education, either exclusively or in addition to internal assessment. It provides a range of detail on the purposes of these assessments and there are several noteworthy findings.

First, the age at which compulsory education ends is shown for each RHPJ. This age matches the typical age of external assessment after basic secondary education in only 6 of the 14 jurisdictions. These are: Japan, Chinese Taipei, Estonia, China, the Russian Federation and Northern Ireland. Clearly, the purposes of the assessment after basic secondary education tend to go well beyond ensuring everyone enters the workforce with a qualification. Related to this finding, Table 3 reveals that all RHPJs using external assessment after basic secondary

education use it again after upper general secondary education, typically at around age 18. These jurisdictions do not view external assessment as something for students to undertake once and once only in their school careers. In this respect, England's typical route through general education of externally assessed GCSEs followed by externally assessed A Levels is in fact the norm among repeat high performers.

Table 3. Repeatedly High Performing Jurisdictions *with* external assessment (national or regional) at the end of basic secondary education

Jurisdiction	Inclusions in the 'top 10s' of seven international comparisons	Compulsory education			Assessment at the end of basic secondary education			Transition from basic secondary to upper secondary education				Assessment at the end of upper general secondary education		
		Duration in years: full-time	Ending age: full-time	Ending age: additional part-time	Typical age	Types	Stated purposes ⁺	Change of institution	Change of general ed. programme	Main age(s) for funnelling /selection of subjects in general education	Typical age for starting vocational secondary education	Typical age	Types	Stated purposes ⁺
Singapore	7/7	9	15	n/a	16/17	National exams	Prog, Qual	Possibly	Almost always	16/17	16/17	18/19	National exams	Prog, Qual
South Korea	6/7	9	15	n/a	15 16	Internal assessment National testing	Prog, Comp Form-nat, Other	Almost always (age 15)	Almost always	16	15	18	National exams	Acc, Prog, Qual, Form-sch, Form-nat
Ireland	5/7	10	16	n/a	15	Internal assessment & national exams	Prog, Qual	No	Almost always	15/16	15/16	17/18	National exams	Prog, Qual
Japan	5/7	9	15	n/a	15 15	National testing Internal assessment & national entrance exams for upper secondary school	Form-sch, Form-nat Comp, Prog,	Possibly	Possibly	18	15	18	Internal assessment & national entrance exams for HE	Prog
Taiwan (Chinese Taipei)	5/7	9	15	n/a	15	National entrance exams for senior high school	Prog	Almost always	Almost always	15	15	18	National exams*	Prog
Poland	4/7	9	15	18	15 15	National exams Certificate	Prog, Qual Prog, Comp	Almost always	Almost always	15	15	19 19	Internal assessment & national exams Certificate	Prog, Qual Prog
Estonia	3/7	9	16	n/a	16	National exams	Acc, Prog, Comp, Form-sch, Form-nat	Almost always	Almost always	16	16	19	Internal assessment & national exams	Acc, Prog, Comp, Form-sch, Form-nat
Belgium: French	3/7	11	16	18	16 All	National exams Internal assessment	Prog Form-sch	No	No	n/a	14	18	Internal assessment & national exams	Prog, Qual, Comp
Canada (e.g. Ontario)	3/7	12	18	n/a	15&16 All	Provincial exams Internal assessment	Prog, Comp Form-sch	No	No	n/a	17	18	Internal assessment & national exams	Prog, Comp

China: B-S-J-Z	3/7	9	15	n/a	15 15	Local authority-administered exams Certificate	Prog, Acc Comp	Almost always	Almost always	15/16	15	18 18	Local authority-administered entrance exams for HE Local authority-administered exams	Prog Comp
Russian Federation	3/7	9/11 [%]	15/17 [%]	n/a	15	National exams	Prog, Comp	Possibly	Possibly	15	15	17	Internal assessment & local authority-administered exams	Prog, Comp
Australia (e.g. NSW)	2/7	11	17	n/a	16	Internal assessment & state exams	Prog, Qual, Comp	No	Possibly	16	14	18	Internal assessment & state exams	Prog, Qual, Comp, Other
England	2/7	11	16	18	16	National exams	Acc, Prog, Qual	Possibly	Almost always	16	16	18	National exams	Acc, Prog, Qual
Northern Ireland	2/7	12	16	n/a	16	National exams	Acc, Prog, Qual	Possibly	Almost always	16	16	18	National exams	Acc, Prog, Qual

Notes: *Abbreviations of stated purposes: Formal accountability (Acc), Completion certification (Comp), Progression (Prog), Qualification (Qual), Informing teaching within stage at school level (Form-sch), Informing teaching within stage at national level (Form-nat), Other student or system purpose (Other).

*A small quota of students from remote areas can combine exam scores with grade point averages (GPAs) to get into university.

[%]Varies across local jurisdictions.

Secondly, in all but one of the 14 RHPJs in Table 3, external assessment after basic secondary education is used for progression purposes. (South Korea is the exception in that it uses teacher assessment instead; its (external) national tests are used to inform teaching within basic secondary education at a national level.) Progression takes on a variety of forms as students transition from basic to upper secondary education. In five of the RHPJs in Table 3, students almost always change institution at this point in their education. In a further five of the RHPJs, including England and Northern Ireland, changing institution is a possibility. It is evident that, excepting South Korea, an important purpose of the external assessments is to guide or even determine these institutional transitions. In Japan and Chinese Taipei, for example, national entrance exams for upper secondary education are held at age 15.

It is also evident that the age at which students transition from one general educational programme to another (for example, from GCSE courses to A Level courses in England and Northern Ireland) often aligns with the age of external assessment. In eight of the RHPJs in Table 3, students almost always change programme after external assessment, and in an additional three RHPJs, students possibly change programme. It is no coincidence that for students progressing to upper general secondary education, the main funnelling or selection of subjects also occurs at the same age as external assessment in 11 of the RHPJs. Furthermore, the typical age for starting vocational secondary education aligns with this age in 10 of the RHPJs in Table 3. When these findings are taken together, they provide a very strong indication that in most RHPJs, external assessment plays a critical role in determining students' directions in upper secondary education. Additionally, it provides students with a final score or grade in subjects that they might never again study.

Among the seven RHPJs that do *not* use external assessment at the end of basic secondary education, we found that the forms of internal assessment used instead vary considerably. That is, no single form of internal assessment is associated with repeat high performers in international comparisons. In Macao, for example, final examinations are organised within schools. In the Netherlands, in contrast, continuous assessment takes place several times per year, with parents receiving progress reports from most schools. This happens in all school years and the final year of basic secondary education is no different. In Lithuania, the approach to internal assessment is different yet again. Each school develops its own procedure for assessing the achievements and progress of its students; the principal approves it and publishes it on the school's website.

Table 4 relates specifically to the seven RHPJs that do not use external assessment at the end of basic secondary education. It provides a range of detail on the purposes of the internal assessments used instead. The key point to note is that, like external assessment, almost any form of internal assessment can have high stakes associated with it for students. In four of the seven RHPJs in Table 4, progression is a stated purpose of assessment at or near the end of basic secondary education. In Finland and Macao, students almost always change institution at the end of basic secondary education, and in Hong Kong, Lithuania and the Netherlands there is the possibility to do so. In four of the seven RHPJs in Table 4, students almost always change programme at the end of basic secondary education, and in one, Hong Kong, students possibly change programme. The main funnelling or selection of subjects in general secondary education also occurs at this point in all RHPJs in Table 4 except the two Belgian jurisdictions. In Hong Kong, Finland and Macau, vocational education pathways begin after basic secondary education.

Not all institutions and courses in upper secondary education have academic entrance requirements. However, performance in assessments, either internal or external, will affect perceptions of ability and potential among teachers, parents and the students themselves.

These perceptions will in turn affect students' confidence and ambitions, influencing the educational choices that they make.

Table 4. Repeatedly High Performing Jurisdictions *without* external assessment (national or regional) at the end of lower/main secondary education

Jurisdiction	Inclusions in the 'top 10s' of seven international comparisons	Compulsory education			Assessment at the end of lower/main secondary education			Transition from lower/main secondary to upper secondary education				Assessment at the end of upper secondary education		
		Duration in years: full-time	Ending age: full-time	Ending age: additional part-time	Typical age	Types	Stated purposes ⁺	Change of institution	Change of general education programme	Main age(s) for funnelling /selection of subjects in general education	Typical age for starting vocational secondary education	Typical age	Types	Stated purposes ⁺
Hong Kong	6/7	9	15	n/a	15	Internal assessment	Prog, Form-sch	Possibly	Possibly	15	15	18	Internal assessment & national exams	Prog, Qual, Other
Finland	5/7	10	16	n/a	16	Internal assessment	Prog, Comp	Almost always	Almost always	16	16	19	Internal assessment & national exams	Prog, Comp
Belgium: German	3/7	11	16	18	All 15/16	Internal assessment Certificate: no info available on assessment	Form-sch Comp	No	No	n/a	14	18	Cert: No further info	Prog, Comp
Belgium: Flemish	3/7	11	16	18	All	Internal assessment	Form-sch	No	No	n/a	14	18	Internal assessment	Prog, Comp
Macao (China)	3/7	10	15**	n/a	15	Internal assessment	Prog, Comp, Form-sch	Almost always	Almost always	15	15	17/18	Internal assessment University-organised entrance exams for HE	Prog, Qual, Comp, Form-sch, Prog
Lithuania	2/7	10	16	n/a	17	Internal assessment	Prog, Comp	Possibly	Almost always	17	14	19 19	Internal assessment Internal assessment/ School or state exams	Acc, Comp, Form-sch, Form-nat, Prog, Qual
Netherlands	2/7	11	16	18	All	Internal assessment	Form-sch	Possibly	Almost always	15	16	17/18	Internal assessment & national exams	Prog, Qual

Notes: ⁺ Abbreviations of stated purposes: Formal accountability (Acc), Completion certification (Comp), Progression (Prog), Qualification (Qual), Informing teaching within stage at school level (Form-sch), Informing teaching within stage at national level (Form-nat), Other student or system purpose (Other).

**Age 15 or the student has completed and passed junior secondary education, up to a maximum of age 18.

An early purpose of external assessment in secondary education: a meritocratic education system

Perhaps the most striking finding from our analysis is that in the majority of the world's most successful education systems, external assessment plays a critical role in determining students' directions in upper secondary education. To many readers, the question of *why* will arise instinctively. Although it is not a fashionable place to look, the answer lies at least in part in the history of examinations.

Meritocracies are social systems, societies, or organisations in which people get success or power because of their abilities, not because of their money or social position (Cambridge University Press, 2021). In 19th century England, a meritocratic education system became a growing aspiration among many educationalists and parents since it could engender social mobility. In the early part of the century, the universities of Oxford and Cambridge, which were the only two degree-awarding institutions in the country, had admissions processes that were more socially than academically selective (Kingdon, 1991). The universities were closely entwined with the upper and professional classes, and a letter from a public school house master was often sufficient for entry. It is highly unlikely that these early 'teacher assessments' were standardised in any way. Their rigour and fairness were highly dubious.

Challenging the status quo, the University of London was founded in 1836 and introduced a university matriculation examination just two years later. A major factor in this was:

"...the desire of the rising middle class to establish a system of higher education that was free of the privilege inherent in the Oxbridge systems."

Kingdon, 1991, p. 33

The author goes on to explain:

"The development of examinations for school students during most of the 19th century [was] inextricably linked with the spread of the University of London degree and pre-degree examinations throughout Britain and the developing British Empire."

Kingdon, 1991, p. 33–34

The aforementioned pre-degree examinations were the progenitors of A Levels.

By around 1850 each university had its own matriculation examinations, and from the 1840s onwards, examinations were used increasingly to support other meritocratic goals. These included recruitment to the civil service, where open competition was encouraged over private favour in order to raise standards. Examinations were also introduced for teacher accreditation (for example by the College of Preceptors), and by the Society of Arts, whose mission was to alleviate poverty through employment (Hudson & Luckhurst, 1954).

At this time there was neither compulsory education nor national provision of education in the country.⁵ However, many aspirational parents wanted regulated schools to allow their children to

⁵ It was not until 1880 that compulsory education existed in England. In this year, the Elementary Education Act made school compulsory from age 5 to age 10. The compulsory education leaving age was

enter into professions such as the army, the clergy, the law and surgery, as well as respectable trades. In 1853 the Society of Arts (which later became the RSA⁶) teamed up with the Mechanics Institute (which offered practical education to the children of artisans and small tradesmen) to offer a 'scheme of examining' (Hudson & Luckhurst, 1954). The scheme included arithmetic, algebra and geometry. Due to short notice and a lack of publicity only one candidate entered (a chimney sweep) and the scheme was cancelled. It ran successfully in 1855 however, and the format of grouped certificate examinations was subsequently adopted by the universities.

Two years later came the founding of the University of Oxford Delegacy of Local Examinations and the University of Cambridge Local Examinations Syndicate (UCLES, now known as Cambridge Assessment) (Watts, 2008). Interestingly, these examination boards were established to set school leaving examinations for scholars who were *not* intending to go to university. The goal was to give both teachers and pupils an achievable aim in school education. From the beginning, emphasis was placed on the importance of consultation with teachers during the construction of examinations, and the boards provided examinations in centres that were local⁷ to candidates (*ibid.*). Both these elements of external assessments such as GCSEs and A Levels continue to this day.

Whilst we have taken England as an example in exploring the introduction of greater meritocracy through examinations, it must be acknowledged that meritocratic principles are axiomatic in the educational systems of all other RHPJs. When compared with China's civil service, for example, England's civil service was a very late adopter of examinations. Civil service examinations were administered in imperial China from 650 CE to 1905, making it the world's longest-lasting meritocracy.

A modern purpose of external assessment in secondary education: accountability

Few educationalists and policy-makers would argue against the centuries-old ideal of progression through merit.⁸ Accountability, in contrast, is often perceived as a more modern and controversial purpose of educational assessments. To build an effective accountability system, it is crucial to balance the various education demands of local stakeholders with the pursuit of the overarching goals of education systems that are assumed to reflect true social priorities (OECD, 2019). This balance can be difficult to strike.

raised to age 11 in 1893, to age 12 in 1899, to age 14 in 1900, to age 15 in 1944, to 16 in 1972 and to 18 (part-time) in 2007.

⁶ The Royal Society for the Encouragement of Arts, Manufactures and Commerce (RSA) was founded in 1754 but its first examinations were not held until 1856. According to the RSA Charter of Incorporation in 1847, its mission was 'The employment of the poor, the increase of trade and the riches and honour of the kingdom' in that order (Hudson & Luckhurst, 1954). The RSA became a part of what is now the OCR awarding body in 1998.

⁷ Hence, 'local' in the names of University of Oxford Delegacy of Local Examinations and the University of Cambridge Local Examinations Syndicate.

⁸ Although when Michael Young coined the term 'meritocracy' in 1958 with his book, *The rise of the meritocracy*, it was in a negative sense.

In many countries, school-level achievement data is tracked over time by administrative authorities (OECD, 2011). Some do this through national testing programmes in which samples or entire cohorts of schools participate. Others make use of assessments which have other primary purposes. In its secondary education system, England focuses on the latter approach. The main indicators of secondary school performance are known as Attainment 8 (a student performance measure) and Progress 8 (a value-added measure), both of which utilise GCSE results.⁹ Headline secondary accountability measures based on Progress 8 are published for national and local authority level and at school level on a website of school and college performance tables, for all to access (Department for Education, 2020).

Tables 3 and 4 indicate that this additional purpose of assessments at the end of basic secondary education may be relatively unusual among RHPJs. That is, in the sources analysed in this study, we rarely found accountability to be stated explicitly as a purpose of assessments at the end of basic secondary education.¹⁰ Of course, this is not to suggest that most RHPJs reject assessment for accountability purposes completely; far from it. International comparisons of students' performances such as PISA, PIRLS and TIMSS yield a degree of accountability at a national level since their results are widely reported and are often used in both public and specialised debate (OECD, 2011). In this sense, by definition, no RHPJs are against assessment for accountability purposes. All use it, albeit not necessarily at the end of basic secondary education. Perhaps the key point to note is that the locus of control of assessment systems differs across different nations, with more or less 'arm's-length control' by different governments in different national settings.

In addition to performances in international comparisons and England's Progress 8 measure, there exist many other approaches to using assessment for accountability purposes. Education systems are complex, with many control factors to be manipulated (Cambridge Assessment, 2017). In 2011, the OECD reported a trend in its countries towards:

"...multi-layered, coherent assessment systems, from classrooms to schools to regional to national to international levels, that: support improvement of learning at all levels of the system; are increasingly performance based; add value for teaching and learning by providing information that can be acted on by students, teachers, and administrators; and are part of a comprehensive and well-aligned learning system that includes syllabi, associated instructional materials, matching exams, professional scoring and teacher training."

OECD, 2011, p. 51–52

It is likely that this trend continues. To give a flavour of the variety and extent of accountability approaches that utilise assessment, we outline what happens in three diverse RHPJs: Finland, Estonia and Shanghai.

Accountability in Finland

Accountability in Finland is much misunderstood by people looking into the country from the outside. 'Felt accountability' is high but assumes a different form to systems which engage in national data collection on every child and school. There is a Finnish survey of student attainment, but this is a

⁹ The Progress 8 score is based on a pupil's performance score across eight subjects; this performance score is known as the 'Attainment 8' score. Attainment 8 takes the average of a pupil's points across a set of their best eight subjects at Key Stage 4 (GCSEs and other general qualifications). A school's Attainment 8 score is the mean of its pupils' Attainment 8 scores. Similarly, a school's Progress 8 score is the mean of its pupils' Progress 8 scores.

¹⁰ Note that absence of evidence is not always the same as evidence of absence, however. Our findings should be interpreted with the caveat that information on accountability was hard to locate.

sample model valid only at the national level, not at school level. Quality in teaching is front-ended through very stringent selection to teacher training and high levels of initial professional qualification (Hancock, 2011). This encourages more convergent behaviour among teachers as an enduring professional expectation, requiring fewer continuing formal accountability measures.

Use of formal tests is high in primary schools, but the results are not reported to the state; instead, results are used to identify children who are at risk of falling behind expectations. Schools have substantial information on student performance (OECD, 2011). If parents complain to municipal administrations, schools will be contacted, can be inspected, and the substantial information on student performance called for and scrutinised. Accountability also assumes a very important, persistent cultural form in Finland. Prior to the massive programme of school closures in Finland for economic reasons (over 2000 schools closed in the past three decades) accountability was assured through the teacher being tightly embedded in village and small-town community – complaints and compliments were direct, and tenure of teachers was long, encouraging responsiveness to parents and community (Autti & Hyry-Beihammer, 2014). Standards now are declining in Finland, and it appears that this culturally reproduced ‘felt accountability’ – although persistent – may not be surviving the structural changes in the school systems and societal and political changes (Heller Sahlgren, 2015).

Accountability in Estonia

Estonia’s educational system differs from neighbouring Finland’s in some ways but is similar in others. This is perhaps unsurprising, given that experts from Finland advised Estonia on education reforms in the 1990s. Student performance is assessed using national examinations, sample-based national tests and regular classroom assessments (European Commission, 2020). The Basic Schools and Upper Secondary Schools Act 2010 establishes external evaluation of learning outcomes; that is, state-level evaluation of the learning outcomes defined in Estonia’s national curriculum. The external evaluation is carried out through all three forms of assessment.

In contrast to Finland, uniform final examinations are held at the end of basic school, at age 16 (Ministry of Education & Research, 2021). The country has embraced the idea that these examinations can serve multiple functions simultaneously, including accountability, and transparency around this position is high. Legislation setting out the objectives of the basic school final examinations sets out their five purposes as follows:

“The purpose of conducting basic school final examinations is to assess the acquisition of general competencies, field competencies, cross-curricular subjects and learning outcomes of the third school level (hereinafter curriculum objectives and learning outcomes of the third school level) in order to:

- 1) provide students, parents, schools, as objective and comparable feedback as possible to the school administrator and the state on the effectiveness of learning and teaching and on the school’s contribution to student progress;*
- 2) explain how the effectiveness of learning and the contribution of the school to the progress of students has changed over time;*
- 3) provide information to the state for making educational policy decisions;*
- 4) support the implementation of the national curriculum and guide the study process through the content and form of the examination;*
- 5) make a decision regarding the graduation of a student from basic school.”*

Ministry of Education and Research, 2015, Section 9.1

Coupled with this regulatory accountability through performance data, formal appraisal of teachers is near-universal and elaborate; on average almost six different methods are used in each school (OECD, 2020). Furthermore, the 'felt accountability' that has been so strong historically in Finland is also likely to be high in this small Baltic nation. Following independence in 1991, Estonia decentralised its school system. This gave schools greater autonomy, including the freedom to make decisions about hiring and dismissing teachers (Schleicher, 2018). Parents now have the right to choose a school for their children, and consequently, schools are competing to attract students. This is at a time when the population of school-age children has declined markedly in recent years; Estonia has some of the smallest secondary school classes in the developed world (*ibid.*).

Accountability in Shanghai, China

Shanghai's regional education system is highly successful by both Chinese and international standards. Although it is not representative of China; Shanghai's population of around 27 million is larger than those of many other countries. The education system is characterised by a highly competitive culture in which much time and resource is invested in teachers and students alike (Schleicher, 2018). It is also characterised by a strong culture of accountability across at least three levels: the classroom, the school and the system.

Whilst teacher appraisal is often led by school principals or school management teams in Shanghai, appraisal by peer teachers is also very common (OECD, 2020). Nearly half of all teachers in the jurisdiction observe other teachers' classes and provide feedback at least once a month. As in Estonia, around six different methods of appraisal are used in each school. The results are used as key criteria for teacher promotions, and Shanghai's well-structured professional career ladder is a powerful incentive to augment teachers' accountability (*ibid.*). In recent years, changes to legislation have ended teachers' automatic lifelong tenure in Shanghai. All basic education teachers must now renew their teacher certification and be evaluated once every five years; five consecutive renewals are needed to obtain tenure (Liang, Kidwai and Zhang, 2016).

At the level of the school, the Chinese education system relies mainly on the regulatory approach to accountability. The Ministry of Education has published a set of school management standards for compulsory education (Ministry of Education, China, 2017) which include goals to promote students' well-rounded development (important in a highly competitive academic culture) and to improve the quality of teaching and learning.

In addition to school inspections mandated by central government, performance-based accountability is used increasingly at this level (OECD, 2020). Across most of China, there are two types of examination at age 15. The 'Zhongkao' (中考) is the highly competitive entrance examination for senior high schools and its key function is to differentiate students' ability. The 'Huikao' (会考) on the other hand, is the graduation examination at the end of junior high school (Zhao, 2021). As compulsory education lasts for nine years in China, junior high schools are held accountable for Huikao examination results, and for their associated completion rates for Year 9 (age 15). There is a national objective for every child in China to complete nine years of compulsory education. This means that the completion rate is one of the key performance indicators for schools and local education authorities. Although these two exams were separate for many years in Shanghai, the Huikao has recently been combined with the Zhongkao there. This means that in Shanghai, the Zhongkao now has the function of accountability at the school and system levels,

although the term Zhongkao is associated nationally with the purpose of progression (*ibid.*). As Liang *et al.* (2016) explain:

“Data from the Zhongkao is used to help policy makers track trends in student learning outcomes. Policy makers also have access to the assessment data disaggregated by district and by school, and each school has student-specific performance scores. The wide scope of information available to policy makers facilitates more effective and targeted use of resources for improvement of education quality.”

Liang et al., 2016, p. 78

Conclusion

The information on the RHPJs collated in this study is transient in nature, existing only as a snapshot in time. Jurisdictions continuously seek to improve their education systems and there is no doubt that some of the details of the status quo reported here will become outdated before too long. What will last, however, is the conclusion that no single approach to assessment is associated with the success of all repeatedly high performing jurisdictions.

Contrary to perceptions expressed in parts of the media, around two-thirds of repeat high performers, including several from across Europe, utilise external assessment at the end of basic secondary education. In many of these jurisdictions it plays critical roles in guiding or even determining students' directions in upper secondary education, and in providing students with qualifications in subjects that they might never again study. This approach is rooted soundly in meritocratic principles. There is no evidence to suggest that alternative approaches are associated with higher student performances at a system level.

A further conclusion is that using assessment at the end of basic secondary education for the secondary purpose of accountability is not necessarily a bad thing, and England is not unique in this respect. Repeat high performers such as Estonia and Shanghai also do so. Our account of Finland confirms that both internal and external assessments can be 'high stakes' for students and for teachers too.

Our conclusions cohere with literature on the need to appreciate the complexity and context of national education systems and to understand relationships across components when introducing change (Mittleton-Kelly, 2003; Cambridge Assessment, 2017). No single innovation will secure a perfect education system and there is no reason to believe that internal assessment is a panacea. Policy-makers should instead focus upon maximising coherence across elements such curriculum, pedagogy and assessment, as this is associated more strongly with successful outcomes.

References

Autti, O. M., & Hyry-Beihammer, E. K. (2014). School closures in rural Finnish communities. *Journal of Research in Rural Education*, 29(1), 1–17.

Baker, K. (2019, February 11). I introduced GCSEs in the 1980s – but now it's time to scrap them. *The Independent*. <https://www.independent.co.uk/voices/gcse-exams-education-skill-shortages-business-schools-robert-halfon-a8774141.html>

Cambridge Assessment (2017). *A Cambridge Approach to improving education*. <https://www.cambridgeassessment.org.uk/Images/cambridge-approach-to-improving-education.pdf>

Cambridge University Press (2021). *Meritocracy*. In Cambridge Dictionary. <https://dictionary.cambridge.org/dictionary/english/meritocracy>

Coe, R., & Heller-Sahlgren, G. (2014). Incentives and ignorance in qualifications, assessment, and accountability. In Heller-Sahlgren, G. (Ed) *Testing worth teaching to – incentivising quality in qualifications and accountability*. The Centre for Market Reform of Education.

Coles, M., & Oates, T. (2009). *Changing qualifications: A review of qualifications policies and practices*. CEDEFOP Report No. 84. https://www.cedefop.europa.eu/files/3059_en.pdf

Crato, N. (2020). Curriculum and educational reforms in Portugal: An analysis on why and how students' knowledge and skills improved. In Reimers, F. (Ed) *Audacious education purposes*. Springer.

Department for Education (2020). *Secondary accountability measures. Guide for maintained secondary schools, academies and free schools*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/872997/Secondary_accountability_measures_guidance_February_2020_3.pdf

Elliot, G., Rushton, N., Darlington, E., & Child, S. (2015). *Are claims that the GCSE is a white elephant red herrings?* Cambridge Assessment research report. Cambridge, UK: Cambridge Assessment.

Elliott, G. (2016). Good – better – best? Identifying highest performing jurisdictions. *Research Matters, A Cambridge Assessment publication*, 22, 37–38.

Elliott, G., Rushton, N., & Ireland, J. (2017). *Is the GCSE in England incongruous in the light of other jurisdictions' approaches to assessment?* Poster presented at the annual conference of the Association for Educational Assessment – Europe. Prague, Czech Republic, 9–11 November 2017.

European Commission (2021). *Eurydice*. <https://eacea.ec.europa.eu/national-policies/eurydice/>

Hancock, L. N. (2011, September). Why are Finland's schools successful? The country's achievements in education have other nations, especially the United States, doing their homework. *Smithsonian Magazine*. <https://www.smithsonianmag.com/innovation/why-are-finlands-schools-successful-49859555/>

Heller Sahlgren, G. (2015). *Real Finnish lessons*. Centre for Policy Studies. <https://www.cps.org.uk/files/reports/original/150410115444-RealFinnishLessonsFULLDRAFTCOVER.pdf>

Hudson, D., & Luckhurst, K. W. (1954). *The Royal Society of the Arts 1754–1954*. John Murray.

Kingdon, M. (1991). *The reform of Advanced A level*. Hodder & Stroughton.

Leaton Gray, S. (2018, August 21). No other European country tests children at 16: Let's scrap pointless GCSEs. *The Guardian*. <https://www.theguardian.com/education/2018/aug/21/no-other-european-country-tests-at-16-scrap-gcses#:~:text=The%20UK%20is%20the%20only,adopting%20a%20more%20enlightened%20a,pproach.&text=Everyone%20learns%20subjects%20that%2C%20in.no%20external%20examinations%20at%2016.>

Liang, X., Kidwai, H., & Zhang, M. (2016). *How Shanghai does it: Insights and lessons from the highest-ranking education system in the world*. Directions in Development. Washington, DC: World Bank. doi:10.1596/978-1-4648-0790-9.

Lough, C. (2020, June 11). State and private schools in 'movement' to scrap GCSEs. *TES*. <https://www.tes.com/news/state-and-private-schools-movement-scrap-gcses>

Ministry of Education and Research, Republic of Estonia (2015). *Conditions and Procedure for the Preparation and Conduct of Level Papers and Basic School and Upper Secondary School Final Examinations and for the Preparation, Assessment and Preservation of Examination Papers*. <https://www.riigiteataja.ee/akt/103072020016>

Ministry of Education and Research, Republic of Estonia (2021). *Uniform basic school final examinations*. <https://www.hm.ee/en/activities/external-evaluation/uniform-basic-school-final-examinations>

Ministry of Education, China (2017). 义务教育学校管理标准 [*Management standards for schools in basic education*].

Mitleton-Kelly, E. (2003). Ten principles of complexity and enabling infrastructures. In Mitleton-Kelly, E. (Ed) *Complex systems and evolutionary perspectives on organisations: the application of complexity theory to organisations*. Elsevier.

Newton, P. (2017). Clarifying the purposes of educational assessment. *Assessment in education: Principles, policy and practice*, 14(2), 149–170.

OECD (2011). *Lessons from PISA for the United States, Strong Performers and Successful Reformers in Education*. OECD Publishing. <http://dx.doi.org/10.1787/9789264096660-en>

OECD (2019). *Strategic Education Governance – Project plan and organisational framework*. <https://www.oecd.org/education/cei/SEG-Project-Plan-org-framework.pdf>

OECD (2020). *Benchmarking the Performance of China's Education System*. OECD Publishing.

OECD (2021). *Education*. <https://www.oecd.org/education/>

Pearson (2014a). *The learning curve. Education and skills for life*. Pearson.

Pearson (2014b). *Pearson Index of Cognitive Skills and Educational Attainment*. <https://www.pearson.com/news-and-research/announcements/2014/05/new-global-educationindexshowsasiansuperpowersexcelinlearning.html>

PIRLS (2016). Progress in International Reading Literacy Study. <https://timssandpirls.bc.edu/pirls-landing.html>

PISA (2018). Programme for International Student Assessment. <https://www.oecd.org/pisa/publications/pisa-2018-results.htm>

Poole, R. (2020). What better time than now to scrap outdated GCSEs for good? *Morning Star*. <https://morningstaronline.co.uk/article/f/what-better-time-now-scrap-outdated-gcse-good>

Rethinking Assessment (2020, October 12). *Open Letter to the Sunday Times explaining the reasons for Rethinking Assessment*. <https://rethinkingassessment.com/rethinking-blogs/open-letter-to-the-sunday-times-explaining-the-reasons-for-rethinking-assessment/>

Richmond, T. (2021). *Re-assessing the future: Part 1 – how to move beyond GCSEs*. EDSK.

Rushton, N., & Elliott, G. (2018). *What lessons from current working practice can be applied to Big Data? Identifying GCSE equivalents across many jurisdictions*. Paper presented at the

annual conference of the International Association for Educational Assessment, Oxford, UK, 9–14 September 2018.

Schleicher, A. (2018). *World Class: How to build a 21st-century school system, Strong Performers and Successful Reformers in Education*. OECD Publishing.
<https://doi.org/10.1787/9789264300002-en>

TIMSS (2019). *Trends in International Mathematics and Science Study*.
<https://timssandpirls.bc.edu/timss-landing.html>

UCAS (2021). *Qualification Information Profiles*. <https://qips.ucas.com/>

UNESCO (2011). *International Standard Classification of Education 2011*.
<http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf>

Watts, A. (2008). Cambridge local examinations 1858–1945. In Raban, S. (Ed) *Examining the World*. Cambridge University Press.

Zhao, J. (2021). Personal correspondence with Regional Director, East Asia, Cambridge Assessment International Education.