Examination Reform:
Impact of Linear and Modular Examinations at GCSE

April 2019

Jo-Anne Baird, Daniel Caro, Victoria Elliott, Yasmine El Masri, Jenni Ingram, Tina Isaacs, Anne Pinot de Moira, Ashmita Randhawa, Gordon Stobart, Michelle Meadows, Caroline Morin, and Rachel Taylor

OUCEA/19/2
Ofqual/19/6506/1

Contents

Acknowledgements

The authors are grateful to Simon Eason (AQA) for providing valuable and continuing guidance for understanding the data characteristics. Additionally, the teachers who participated in the qualitative research provided invaluable insights into the use of linear and modular examinations in schools. We are also grateful to the following people for their contribution to the research process: Carol Brown, Jessica Chan, Therese Hopfenbeck, Steve Holmes and Jane McNicholl.

Glossary of terms and abbreviations

ACME	Advisory Committee on Mathematics Education
ACVE	Advanced Certificate of Vocational Education
AGNVQ	Advanced General National Vocational Qualification
Accelerated learning	This term is used in a variety of ways. It can refer to pupils taking a course over a shorter period of time with fewer contact hours than most others do. Accelerated learning can also denote different pedagogic approaches that aim to enhance pupils' learning or performance. Both interpretations imply an objective of organising teaching and learning to help pupils progress as far and as fast as possible. In the UK, the notion of fast-tracked learning is often twinned with early entry to examinations.
Assessment	The act of judging or deciding the amount, value, quality or importance of something, or the judgment or decision that is made.
BTEC	BTEC formerly stood for Business and Technology Education Council, an awarding body which no longer exists. Nowadays BTEC Nationals are qualifications owned by the awarding organisation, Pearson, used for progression to employment or university.
Cashing in	In a modular examination system it refers to aggregating all the credits earned into an overall grade of a subject.
CBMS	Credit-based modular system
Certification	A system to formally recognise people's capabilities or attainment through candidates following standard assessment procedures. This validating process provides objective and reliable references to outsiders about the knowledge and skills the qualification holders possess. This term is also used for the issuing of examination results.
Controlled assessment	An internal assessment conducted by teachers within schools as a component of their students' GCSE. It was designed to assess some aspects of learning (e.g. planning, collaborative work) that are not assessable by an external, timed written examination. It replaced coursework as an attempt to mitigate concerns about plagiarism and cheating and thus enhance the credibility of GCSE.

Coursework	An internal, school-based examination administered and assessed by teachers to their students sitting for GCSEs. Often coursework is used for evaluation of subject-specific skills not easily assessed by a pen-and-paper examination. Coursework results constitute a percentage of the overall examination grade.
CQFW	Credit and Qualifications Framework for Wales
Early entry	Being entered for an examination at an earlier age than typically scheduled. For example, certifying for GCSE examinations before the end of Year 11, when certification typically takes place.
EQF	European Qualifications Framework
Examination	A formal test of a person's knowledge or proficiency in a subject or skill.
Examination entry	Administrative process involved in putting a candidate forward for an examination. This term can also be used to imply the sitting of an examination by a candidate.
External assessment	Assessment that is external to the school – usually conducted by the examination board. These are typically examinations.
FHEQ	Framework for Higher Education Qualifications for England, Wales and Northern Ireland
Functional Skills	Vocational English and maths qualifications that assess the skills needed for everyday life and work.
Grade boundary	The mark (score) required to be awarded a specific grade (e.g. C). Sometimes called a cut-score.
GCE	General Certificate of Education - examinations were divided into Ordinary Level (O level), designed for 16-year-olds and Advanced Level (A level), for 18-year-olds. GCEs preceded GCSEs.
GCSE	General Certificate of Secondary Education, a qualification taken by pupils aged 14-16 in England, Wales and Northern Ireland.
GNVQ	General National Vocational Qualifications, designed as equivalent of GCSE and A level in vocational education. It was no longer offered after 2007.
International GCSE	Alternative qualifications offered by exam boards to the international and domestic market, includes Cambridge IGCSEs®
Internal assessment	Assessment that is internal to the school and is teacher assessment, typically coursework. Often referred to as 'non-examined assessment'.

Key Stage 3	A three-year curriculum stage in England, ending at Year 9. National examinations in English, mathematics and science at Key Stage 3 were administered until 2008.
Key Stage 4	A two-year curriculum stage in England, ending at Year 11. National examinations including GCSEs are available at Key Stage 4.
Linear examination	Students sit all of their exams in one series at the end of the course of study.
Modular examination	The totality of the assessment is broken into discrete units for assessment, the results of which are combined to give an overall result.
Modular flag	Variable in the dataset distinguishing modular from linear examinations.
NPD	National Pupil Database
NQF	National Qualifications Framework
NVQ	National Vocational Qualification, a work-based qualification which recognises the skills and knowledge a person needs to do a particular job.
QCF	Qualifications and Credit Framework. This has been replaced by the RQF.
Re-sitting	Taking an assessment more than once.
RQF	Regulated Qualifications Framework
Semesterisation	Usually associated with the organisation of the academic year in colleges or universities. In secondary education, this relates to unitising curricula so that they can be delivered in a shorter time.
SES	Socio-economic status
SCQF	Scottish Credit and Qualifications Framework
Terminal assessment	Assessment taken at the end of the course, just prior to certification.
Unitisation	A qualification formed by a number of discrete parts. Each unit can be assessed individually. Unitisation can allow combination awards, involving units from different courses. Unitisation focuses upon awarding of qualifications which are separated into discrete parts, whereas modularisation emphasises dividing the curriculum.
VGCSE	Vocational General Certificate of Secondary Education

Glossary of organisations

AQA	Assessment and Qualifications Alliance http://www.aqa.org.uk/
CCEA	Council for the Curriculum, Examinations & Assessment http://ccea.org.uk/
Cambridge Assessment	http://www.cambridgeassessment.org.uk/
City & Guilds	http://www.cityandguilds.com/
DfE	Department for Education https://www.gov.uk/government/organisations/department-for-education
JCQ	Joint Council for Qualifications, which is an umbrella body for Examination Boards http://www.jcq.org.uk/
Ofqual	The Office of Qualifications and Examinations Regulation, a non-ministerial department that regulates qualifications, examinations and assessments in England. https://www.gov.uk/government/organisations/ofqual
OCR	Oxford, Cambridge and RSA Examinations http://www.ocr.org.uk/
Pearson	http://qualifications.pearson.com/en/home.html
QCA	Qualifications and Curriculum Authority http://www.qca.org.uk/
Qualifications Wales	http://qualificationswales.org/english/
SCAA	School Curriculum and Assessment Authority
SQA	Scottish Qualifications Authority http://www.sqa.org.uk/sqa/70972.html
WJEC	Welsh Joint Education Committee http://www.wjec.co.uk/

Policy Steer from the Secretary of State for Education (2010-2014)

Rt Hon Michael Gove MP
Secretary of State

Sanctuary Buildings Great Smith Street Westminster London SW1P 3BT
tel: 0370 0002288 www.education.gov.uk/contactus

Glenys Stacey
Chief Executive
Office of Qualifications and Examinations Regulation
Spring Place
Coventry Business Park
Herald Avenue
Coventry
CV56UB

6th February 2013

Ofqual policy steer letter: reforming Key Stage 4 qualifications

Thank you for the advice that you have provided on my proposals for reforming Key Stage 4 qualifications, on which I have recently consulted. I am publishing the Government's response to that consultation today, and am now writing to you to set out my policy steers on the development of the new qualifications.

Following our consultation I remain persuaded that there is an urgent need for reform, to ensure that young people have access to qualifications that set expectations that match and exceed those in the highest performing jurisdictions. There was broad agreement in the meetings we held during the consultation period that GCSEs as currently constituted are not giving our pupils the best chance to succeed, and that change is required. However many have also argued convincingly that GCSEs themselves could, with comprehensive reform, once again be highly respected qualifications in which pupils, employers and further and higher education institutions can have faith.

I have therefore decided that GCSEs should be comprehensively reformed, building on the work that Ofqual has already done to strengthen the qualification. The qualification will remain, but will be subject to significant reform in order to command the respect our pupils deserve as reward for their hard work. This letter sets out the scope of the reforms I believe are necessary. I would like to see those changes applied to GCSEs, ready for first teaching by September 2015, in at least the following subjects: English language, English literature, mathematics, biology, chemistry, physics, combined science (double award), history and geography. Other subjects may be in a position to move to the new approach by that date as well. Changes to remaining subjects should follow as soon as possible after that. I believe

our aim should be for that to happen for all subjects for first teaching in September 2016; I would welcome your views on the extent to which that will be possible. I would be grateful if Ofqual would begin work immediately on revising the GCSE regulatory requirements as informed by the policy set out in this letter with a view to giving schools at least a year to prepare for first teaching.

GCSEs will continue to have multiple purposes. The primary purpose is to evidence pupils' achievement against demanding and fulfilling content. They also need to provide a strong foundation for further academic and vocational study; and a basis upon which schools will be held accountable for the performance of all of their pupils. I understand the challenge of creating qualifications that are both reliable and stretching (requiring less predictable assessments and less scaffolding, for example). I am clear that the value of the qualifications for individuals must take precedence ahead of ensuring the absolute reliability of the assessment. We will take account of this in considering the implementation of the new accountability framework.

I have considered carefully the points you have made about the potential risks of reforming the qualifications market at the same time as fundamentally changing the qualifications themselves. I remain very concerned about the perverse incentives in the current system that have led to the 'race to the bottom'. I also recognise the significant steps you have taken to tighten up the regulation of GCSEs and to ensure that standards are right. In this context I am persuaded by your advice that we should not move to a single Awarding Organisation offering each subject suite at this time. I intend to keep this position under review should it become clear that there remains a tension which acts against our shared priority of rigorous and challenging qualification standards.

You have highlighted the interaction between qualifications and the way in which schools are held accountable, identifying the focus the current system places upon pupils near borderlines. I am today launching a consultation on accountability, which sets out how we propose to address this, for example, by using average points score measures alongside a threshold measure. The latter will focus on those subjects – English language and mathematics – in which achieving good qualifications is hugely important for pupils' prospects of continuing to further study and employment. You will want to take proposals for the new accountability system into account when designing your regulatory arrangements for the qualifications.

The reformed GCSEs should remain universal qualifications of about the same size as they are currently, and accessible, with good teaching, to the same proportion of pupils as currently sits GCSE exams at the end of Key Stage 4. At the level of what is widely considered to be a pass (currently indicated by a grade C), there must be an increase in demand, to reflect that of high-performing jurisdictions. This is something we believe the vast majority of children with a good education should be able to achieve. At the top end the new qualification should prepare pupils properly to progress to A levels or other study. This should be achieved through a balance of more challenging subject content and more rigorous assessment structures. We know that employers and others are keen for greater reassurance that pupils who

achieve that level of performance in English and mathematics are literate and numerate.

The qualifications should be linear, with all assessments taken at the end of the course. I am concerned that the current system of tiered papers, whereby pupils are forced to choose between higher and lower tier papers, places a cap on ambition. I would like reformed GCSEs to avoid that, while enabling high quality assessment at all levels. The appropriate approach to assessment will vary between subjects and a range of solutions may come forward, for example, extension papers offering access to higher grades alongside a common core. There should be no disincentive for schools to give an open choice of papers to their pupils.

Reformed GCSEs must prioritise stretching assessment, which truly tests the depth and breadth of pupils' knowledge and abilities. Examinations must test extended writing in subjects such as English and history, have fewer bite-sized and overly structured questions, and in mathematics and science have a greater emphasis on quantitative problem-solving. Internal assessment and the use of exam aids should be kept to a minimum and used only where there is a compelling case to do so, to provide for effective and deep assessment of the specified curriculum content. You will want to consider this in the light of your current review of controlled assessment.

I consider there to be a strong case for the reformed GCSEs to have a new grading scale, to reflect the step change in expectations for pupils, and would welcome your advice on this. Any changes should apply across all subjects, and should differentiate performance more clearly, particularly at the top end. For qualifications in English language and mathematics I would like you to consider the benefits of all pupils receiving more information directly from Awarding Organisations on their performance across the different areas tested by the qualification, in order particularly to support progression for those who may need to re-take the qualification post-16. I recognise that there is a tension between the provision of more detailed marks and the challenge of reliably marking more open-ended assessments, and that you will want to work with Awarding Organisations to determine what is possible here.

We will publish for consultation, by May this year, requirements for subject content ir the new qualifications in English language, English literature, mathematics, science, history and geography in time to allow Awarding Organisations to prepare specifications.

The new GCSEs should include English literature and English language but not a combined 'English' option. They should include a combined science option worth twc GCSEs but not a combined science option worth one GCSE. We are considering what the subject suite should be in mathematics and will confirm in due course. I am keen to see qualifications developed in a wide range of modern and classic languages. We do not anticipate publishing content requirements for subjects outside the EBacc. I would welcome your view on the appropriate subject coverage of reformed GCSEs.

It may be helpful for there to be some form of regular post-assessment review of reformed GCSEs, and I would welcome your views on whether arrangements similar to those proposed for the new A levels could be introduced.

I would like to thank Ofqual for its work to date on reforming qualifications and for your helpful recommendations on the ways in which they should be introduced. I look forward to continuing to work with you, ensuring that all young people are given the best possible chance to succeed and provided with qualifications which are rigorous and relevant, raising the bar on the standards we expect them to achieve.

I am copying this letter to Leighton Andrews AM, John O'Dowd MLA, Graham Stuart MP, and Sir Michael Wilshaw.

MICHAEL GOVE

Chapter 1 The project and its policy context

From 2012, the rules around General Certificate of Secondary Education (GCSE) qualifications, which were designed to allow students to take assessments over the course of study, were modified to require students to take all of the assessments at the end of the course. This change to the rules turned GCSEs designed to be 'modular' in structure into a 'linear' examination structure. More fundamental reforms begun in 2013 cemented this move to linearity, redesigning the syllabuses and examination structure to fully deliver linear qualifications. Reforming the examinations in this way was a response to concerns that modular examinations led to repeated testing that disrupted teaching and were partly responsible for a perception that examination standards had declined in England. In the policy steer letter to Ofqual (reproduced above), from the then Secretary of State for Education, Michael Gove indicated his requirements for curricular and examination reform. The new examinations were to be more demanding and linear in structure.

Examination reform is not typically cumulative, with each reform building upon a previous trajectory. Instead, they are cyclical, with the same reforms or underlying issues recurring (Baird & Hopfenbeck, 2016). In part, this is due to a lack of policy memory on the part of the system (Hodgson & Spours, 2003). Importantly though it is caused by different values being prioritised by changing Ministers of Education, who have the constitutional and legal responsibility for setting broad assessment policy. As it takes a long time to institute examination reforms and electoral cycles are only five years' long in England, reforms are sometimes announced before the previous set has come to fruition, and there is infrequent evaluation of the impact of reforms. Thus, research on the effects of linear and modular examinations is pertinent to current examination policy in England, but it is also important for future policy and for other country contexts. On 14 December 2011, Schools Minister Nick Gibb stated that he wanted to:

> … break the constant treadmill of exams and retakes throughout students' GCSE courses – school shouldn't be a dreary trudge from one test to the next. Sitting and passing modules has become the be-all and end-all, instead of achieving a real, lasting understanding and love of a subject. Students shouldn't be continually cramming to pass the next exam or re-sitting the same test again and again simply to boost their mark – then forgetting it all by moving onto the next module immediately.[1]

The GCSE examinations are high stakes for students and teachers, as the results are used to select students for A level and other courses, as well as in school accountability performance tables. Public confidence in them is important and is currently high (YouGov, 2019). As Ofqual is charged with maintaining this confidence and qualification standards, understanding the effects of examination structure matters for future examination reform planning. Examinations are a

[1] https://www.gov.uk/government/news/end-for-gcse-modules-and-spelling-punctuation-and-grammar-marks-restored-to-exams

large cost to the public purse, with expenditure on examinations having risen in state schools from £153 million in 2002-2003 to £328 million in 2010-2011; triple the rise that inflation alone would have caused (Ofqual, 2015a). Modular examinations at GCSE have been a large part of that rise, as students took more examinations and re-sit modules. Of course, larger student numbers and students taking more GCSEs in general are also factors in the rising examination bill for the country. Profound questions regarding the suitability of GCSE examination structures for student learning, effects upon public confidence and value for money therefore need to be considered.

This project has been conducted collaboratively between Oxford University Centre for Educational Assessment and Ofqual and is funded by Ofqual. The aims of the project were to add to what is known about the effects of linear and modular examination structure by:

- placing GCSE examination structure reforms in context, including in relation to:
    - policy rationales relating to these examination structures,
    - the historical introduction of modular and linear examinations,
    - international research on the effects of modular examinations,
    - subject-specific issues with these examination structures;
- synthesising previous research on the effects of modularisation and linearisation of examinations;
- conducting large-scale analyses on the longitudinal effects of reform upon examination outcomes and attainment gaps;
- investigating current changes to teaching and examination entries resulting from contemporary reforms;
- assessing the impact of the examination structure change on the standard setting process;
- exploring the economic effects of the policy reform.

Our empirical work involves quantitative research on the GCSE outcomes in modular and linear examinations between 2007 and 2014. There is a focus upon GCSE English, mathematics and science outcomes and we drew upon aggregated and pupil-level data. These subjects were amongst the first to be included in the most recent round of reforms. The research in schools involved 84 interviews with teachers of English or mathematics or head teachers. The interviews were conducted prior to the structural changes to the examinations and again following them.

Over the period of this research, a range of policy issues affected schools. In terms of GCSE assessment policy, the demands of the subject content has been raised, the use of tiered examinations was reduced, as was the amount of internal, teacher assessment and the grading scale was changed from letter grades to numbers. Changes were also made so that early and multiple entry for examinations would reduce. A levels were reformed contemporaneously, also affecting schools. As discussed later in the report, there have been changes to the school performance tables that form part of the accountability mechanisms. However, broader issues may have had greater effects than the examination

reforms on examination outcomes and teacher practices. Following the 2008 global economic crisis, the UK government instituted a range of austerity measures. According to the National Audit Office, school budgets have a real-term reduction of eight per cent per pupil between 2014-15 and 2019-20 (National Audit Office, 2016). However, since the research was conducted for this report, there has been an extra £1.3bn to the schools budget over 2018-19 and 2019-20, meaning that funding has been maintained in real terms per pupil across these two years.[2] Teacher shortages have been problematical, with secondary school teacher numbers falling by approximately five per cent between 2010 and 2016 (National Audit Office, 2017). Shortages have been particularly acute in mathematics and sciences subjects, despite the £67m mathematics and physics teacher supply package policy (National Audit Office, 2018a). Teacher practices and school policies will also have been affected by the academisation programme, which had impacted upon one third of schools in England by January 2018 (National Audit Office, 2018b). Social trends such as the rise in mental health problems are also affecting schooling, with a recent estimate indicating that one in ten children have a diagnosable mental disorder (Department for Education (DfE), 2017). Examination stress is likely to interact with these issues. Real-world research entails complexity and the findings must be interpreted in the knowledge that they are caused by a number of factors, not only the change to examination structure.

Evaluating examination policy changes is essential for the country to effect educational progress. Otherwise, policy decisions have to be taken without an evidential basis. Reform is costly and education is key for the knowledge economy so it is important that resources are channelled effectively. Timing of the research is critical if we want to know how the reforms changed practice. In this study we were able to conduct research in schools just before the new syllabuses were implemented and we visited schools again in the first year of the examinations for the new syllabuses. Therefore the research scrutinises teachers' experiences of the reforms as close to when they occurred as possible. Naturally, those effects change as the system becomes accustomed to the reforms. Future studies could address those issues, but the findings of the current research are salutary with respect to the remarkable resilience of the system to reform.

The meaning of modular and linear assessment

Structural definitions of modularity and linearity

In the glossary at the beginning of this report, we define modular assessment as being broken into discrete units, the results of which are combined to give an overall assessment. This operates well enough as a working definition, but it does not distinguish modular from linear assessment completely. After all, if the curriculum is large, it is often the case that there is more than one component to an examination, involving multiple question papers and perhaps coursework, a practical or other performance. Thus, modular assessment has to involve more than the unitising of assessment and typically implies staged assessment. Even

[2] https://www.gov.uk/government/speeches/justine-greening-statement-to-parliament-on-school-funding

so, coursework can be completed throughout a course and so the distinction between modular assessment and coursework is not cut and dried. Various relationships between curricular content and modules have been constructed, including complementary modules that can be taken at any time and in any order, sequential modules in which performance in one module builds upon the knowledge and skills of other modules or articulated modules in which there are formal prerequisites (Warwick, 1987, p.83).

To complicate the definition, in practice in modular GCSEs a large minority of students took all of the examinations at the end of the course, effectively taking a modular examination structure in a linear fashion. Modular examinations typically permit re-sitting and some allow early entry to all of the modules before the anticipated course duration. Linear examinations might be composed of various units, but they are all taken at the end of the course and re-sitting typically means taking the entire set of assessments again.

Many definitions of modularity have been proposed. The key difference between them is in the relationship between the module and the qualification system as a whole (Ertl & Hayward, 2010, p.384). 'Radical' versions of modularisation involve credit-based systems in which qualifications can be compiled from a set of stand-alone components (Ertl, 2001; Pilz, 2002). For GCSE, the definition of modularity is less extreme; modules are building blocks that form part of a larger whole. It is useful to consider three forms of modularised curriculum: internal, external and connective (Table 1.1; Ertl & Hayward, 2010, p.386-7). In the main, GCSE modularisation was 'internal', in which each syllabus was decomposed into constituent units. However, in GCSE science, modules could form double science or separate science subject modules. Additionally, introduction of the Diplomas involved incorporation of GCSE modules and even entire GCSE or functional skills qualifications. Indeed, use of modules across academic and vocational qualifications is one strategy for tackling parity of esteem[3] of the two progression routes (Davies, 1991; Raffe, 1994). To summarise, there were elements of the 'external' definition of GCSE modularisation, but it was mainly an 'internal' approach.

Table 1.1 Forms of modularity

	Internal	External	Connective
Qualifications	Subdivided into modules	Modules can be part of several qualifications	Unified qualification system composed of modules
Credit	Only within particular qualifications	Across specific qualifications	Across entire qualification system
Curriculum	Separate syllabuses	Qualified body specifies ways in which modules can be combined to form qualifications across specific syllabuses	Individualised curriculum for lifelong learning

[3] This term is often used to suggest that vocational (NVQ and GNVQ) and academic (GCSE and GCE A level) qualification routes would be held in equal regard, by parents, employers, and higher education institutions (see http://www.oxfordreference.com/view/10.1093/oi/authority.20110803100306642).

Qualifications frameworks

The Qualifications Credit Framework (QCF),[4] introduced in England in 2008, gave all registered qualifications a level and tariff so that they could be compared in a standardised manner in terms of demand and breadth. This was in keeping with the 'connective' definition of modularity, and the narrative around the QCF in policy circles conformed with the aspirations of connective modularity; that young people would be able to bank credit for their achievements on modules over a lifetime of learning. This learning could be tailored to their specific needs. The QCF has been removed as it was believed to be a potential impediment to the production of valid qualifications. Wider moves to standardise approaches to crediting qualifications for purposes of transparency and transferability across systems internationally coincided with the QCF. The European Qualifications Framework (EQF)[5] acts as a reference point that facilitates understanding of the relative level of qualifications that are taken in different European countries. It remains to be seen whether the UK exit from the EU will affect its use.

Note that within the UK, there are four frameworks which encompass most of the qualifications; the Regulated Qualifications Framework (RQF),[6] the Scottish Credit and Qualifications Framework (SCQF), the Credit and Qualifications Framework for Wales (CQFW) and the Framework for Higher Education Qualifications for England, Wales and Northern Ireland (FHEQ).[7] The Bologna Framework,[8] applied to higher education, involves a European Credit Transfer and Accumulation System[9], which aims to recognise quality-assured student achievement across Europe, thereby facilitating student mobility. Credit frameworks can be communicative devices, used for reform of qualification systems or exercised more radically as transformational devices (Raffe, 2011). To summarise, there are international drivers for connective modularisation with high-level aims. Whilst these frameworks are coherent at an administrative level, full-blown connective modularisation can produce incoherent curricula. Additionally, the flexibility that is a desirable feature of these credit frameworks does not in itself remove structural inequalities in education systems (Raffe, 2009), even those associated with academic and vocational parity of esteem. Next, we turn to the different meanings associated with modular and linear examination results for individual students.

Score meaning

By assessing students in a linear or modular fashion, we gauge their learning differently; literally measuring differently. Education has the, oftentimes conflicting, goals of emancipation and selection. Modular examination structures could be seen as better aligned with emancipatory goals, as they are built upon the notion of effort leading to results. Linear examinations could be seen as

[4] https://www.gov.uk/what-different-qualification-levels-mean/compare-different-qualification-levels Note that Ofqual launched a consultation on a new QCF on 25 March 2015. https://www.gov.uk/government/consultations/after-the-qcf-a-new-qualifications-framework

[5] http://ec.europa.eu/ploteus/search/site?f%5B0%5D=im_field_entity_type%3A97

[6] https://www.gov.uk/government/consultations/after-the-qcf-a-new-qualifications-framework

[7] http://scqf.org.uk/content/files/europe/QFUK_Joint_Report_-_Updated_March_2010.pdf

[8] http://www.eua.be/eua-work-and-policy-area/building-the-european-higher-education-area/bologna-basics/Bologna-an-overview-of-the-main-elements.aspx

[9] https://ec.europa.eu/education/resources/european-credit-transfer-accumulation-system_en

better aligned with selection goals and, indeed, commentators often conflate linear, high-stakes examination results with intelligence.

Aggregated scores over a staged assessment give an average of a student's performances at different time-points, whilst they are learning the curriculum. Thus, staged assessment outcomes arguably under-weight the culmination of the student's learning: the point that they have reached at the end of the course. To do well on a modular course, students must apply themselves throughout. Equally, as re-sitting opportunities are often a feature of modular assessments, the examinations must be highly reliable or students' scores could be improved by chance. Further, re-sitting policies are compatible with interpretation of scores as representations of effort as well as ability because they allow students to try hard and make another attempt. Due to the decomposition of the curriculum into smaller units and the availability of many examples of modular question papers, modular assessment scores are sometimes associated with drilling for short-term goals on superficial content. Arguments against the validity of modular examinations focus upon students' apparent lack of breadth and depth of understanding of the subject gained from their studies.

A linear examination score represents the performance a student gave at the end of a course and is therefore reliant upon the student's memory of material taught and learned across the duration of study. Cramming for finals is an aspect of preparation for linear examinations that affects score interpretation, as people with good examination technique and memory capacity can do well under these conditions. Bright, but lazy, students might do well under a linear system. Without the disruption of frequent testing, linear examination scores could potentially represent more in-depth and consolidated learning, with students given the opportunity to study broadly, make connections across the materials to which they have been exposed and to understand the structure of the subject. With fewer assessment opportunities, the examinations must be reliable to represent the students' attainment fairly. Arguments against the validity of linear examinations tend to focus upon the unfair nature of a high-stakes, all-or-nothing assessment, as it might not well represent students' actual levels of knowledge and skills and could be affected by the stress of examinations, amongst other short-term factors. The meanings attached to the results from the two examination structures underlie the policy changes that we describe next.

GCSE examination structure policy

To explain the policy of GCSE examination structure, we need to go back to the period before the examinations were introduced. The General Certification of Education (GCE) was introduced in 1951, replacing the School Certificate and Higher School Certificate. The examinations were divided into Ordinary Level (O level), designed for 16-year-olds and Advanced Level (A level), for 18-year-olds. O levels were intended for the top 20 per cent of students; Certificates of Secondary Education (CSE) were introduced in 1965 to cater for the next 40 per cent of 14- to 16-year-olds. In 1988 the General Certificate of Secondary Education (GCSE), which catered to an even wider ability range, replaced both O levels and CSEs (Daugherty, 1995). An examination that caters for almost the

entire ability range faces the twin challenges of being at once appropriate for the test-takers and distinguishing between their performances sufficiently.

GCSEs combined content and standards from O levels and CSEs and also met the needs of the bottom 40 per cent. It was a daunting task to assess adequately both the most and least able in one qualification. The new qualifications were underpinned by qualification and subject specific criteria; most subjects offered two or more tiers of assessment, each aimed at part of the cohort, and each based on positive achievement (Torrance, 2002; Daugherty, 1995; Lambert & Lines, 2000). The basic tenet of GCSE was to motivate all students to achieve their best performance (Meyer, 2011). Until 2013, the school leaving age in England was 16 and the GCSE was the school-leaving examination. Almost the entire cohort at age 16 takes at least some GCSEs, especially in English, mathematics and the sciences. However, young people now have to stay in education or training until they are 18, so the GCSE is no longer a school-leaving examination. With their use in school evaluation too (of which more later), they are still a central feature of the education system.

Ofqual regulates GCSEs through statutory rules and guidance. The subject content is owned by government and Ofqual must confirm that it can be regulated effectively (that it supports valid qualifications) before adopting it into its regulation. Until recently regulation was conducted through subject criteria and codes of practice, the latest of which was published in 2011 (Ofqual, 2011). These have now been replaced by the General Conditions of Recognition, Qualification and Subject level Conditions and Guidance. Major, relevant policy changes to the GCSEs are given in Appendix A. The vast majority of GCSEs until 2009 were linear in structure; that is, assessed at the end of the two-year course. Subject criteria set out the knowledge, understanding, skills and assessment objectives common to all syllabuses in that subject. Subject criteria were intended to help ensure consistency and comparability of standards across the examination boards. However, secondary school teachers complained that they had to teach to the examinations and that reading around a subject was no longer rewarded (Isaacs, 2010; Isaacs, 2012). Ofqual's rules and guidance aim to ensure quality, consistency, accuracy and fairness in the assessment and awarding of qualifications, as well as to help maintain standards across syllabuses both within and between examination boards, and from year to year (Ofqual, 2009; Ofqual, 2011). Through these devices, examination structure policy is formalised.

When GCSEs were introduced in 1988, there were very few modular examinations and those were all in science. Prior to 2009 only a small number of subjects such as the sciences, mathematics and modern foreign languages, were available in modular format. However, not all examination boards offered a modular format and those that did generally offered the same subjects in linear fashion. GCSE qualification and subject criteria were non-restrictive about modularisation since they were first promulgated, but never required it.

Following the Dearing Review (1996), unitisation became mandatory for A levels starting with Curriculum 2000. The Dearing Review promoted a unit approach to

lend parity of esteem to vocational provision (1996, p.13) such as GNVQs (General National Vocational Qualification) and Advanced Certificates in Vocational Education (ACVE), both of which became Applied A levels. Reflecting these broader changes in policy towards modular examinations, the 2004 *QCA Code of Practice* recognised that some GCSEs were already modular and stated that units could only be re-sat once, with the better mark counting toward the final grade (Qualifications and Curriculum Authority (QCA), 2004). Vocational GCSEs (vGCSE), which were modular, had been introduced in September 2002. The same year's criteria for accreditation of external qualifications exempted traditional GCSEs from its rules for unitisation on the grounds that the examination had not historically been unitised (QCA, 2004). After the Dearing Review, modularisation of GCSE thus first arose in vocational subjects, with new qualifications being introduced (vGCSE) to create vocational pathways with declared parity of esteem with academic progression routes.

GCSEs were re-evaluated in 2007 and re-developed for first teaching in 2009. However, in 2007 most of the policy attention was upon the Diploma qualifications. Sir Mike Tomlinson's (2004) review of 14-19 qualifications had proposed a Diploma system of assessment to replace GCSE and A level (DfES, 2005), sharing Dearing's (1996) aim of creating parity of esteem between qualification pathways. The policy response from the government was to retain the academic A levels and GCSEs. The Diplomas developed in 2007 were solely in vocational subjects. Importantly, Diplomas were unitised. They also incorporated a large proportion of teacher assessment. Notwithstanding the Diploma policy drivers, the re-development of GCSEs took place during a time of public scepticism about coursework (which is teacher-assessed in England) and standards of examinations (QCA, 2005). Clearly, there were conflicting policy narratives at work in different areas of assessment policy.

So, in 2007, QCA wanted the new GCSEs to complement the new Diploma qualifications that were being promoted in policy circles as a revolutionary qualification that would change the face of the system (Isaacs, 2013b). Another driver for standardisation of approach was the QCF, which was designed to bring clarity to the equivalence of the plethora of qualifications on offer to young people and schools. Although not documented formally, policy makers were certainly discussing the possibility of aggregated qualifications that could incorporate modules from current vocational and academic qualifications. Indeed, the Diplomas had already made this a possibility and a new national computer system, the Diploma Aggregation Service (DAS), which collated results from qualifications across examination boards, was needed to operate the grading of the Diplomas.

In an internal meeting paper on revisions to the GCSE criteria addressed to the QCA executive committee in June 2007, Isaacs and Curnock Cook, wrote that 'in general terms, we are keen to promote unitisation of GCSEs' (unpublished QCA executive meeting decision paper, 4 June 2007, p.3). Notwithstanding the policy paper, Isaacs raised reservations about this in internal discussions, primarily because students would be overburdened by assessment. However, there was a drive at the time to align as many qualifications as possible with the new

Diploma qualifications, which were completely unitised (Isaacs, 2013b). The executive committee paper warned that with unitised qualifications, assessment could become fragmented and that students might re-sit modules in order to gain a grade C, 'especially for schools wanting to push their league table positions to the maximum' (p.4). The solution proffered was to put in place a one re-sit rule and to have a minimum of 40 per cent of the assessment at the end of the course (Isaacs, 2014; Ofqual, 2012; QCA, 2007a).

There was a public consultation in the summer of 2007 on the new subject and qualifications criteria; 64 per cent of respondents believed that the GCSE qualification criteria were appropriate, 43 per cent did not believe that the unitisation of their GCSE subject would cause any problems. However, 35 per cent thought it would. A majority of respondents (71%) agreed that at least half of all assessment should be at the end of the course, to ensure that assessment in unitised GCSEs did not become fragmented and atomised (QCA, 2007b; Vidal Rodeiro & Nádas, 2011).

Unexpectedly, the examination boards submitted GCSE draft qualifications for accreditation in 2007 that were almost exclusively modular in structure. Examination boards clearly considered that modular examinations would be more appealing to teachers. In response, the 2008 GCSE qualifications criteria were restricted to preclude too many different modular structures. The new rules stated that unitised GCSEs must:

- contain a maximum of four assessment units in a single award
- allocate a weighting of at least 20% to each assessment unit
- allow only one re-sit of an assessment unit with the better result counting towards the qualification
- allocate a weighting of at least 40% to terminal assessment

(QCA, 2008; Rodeiro & Nádas, 2011)

Surprisingly, the modularisation of GCSEs was absent in policy documents – a QCA publication from 2008 on the changes to GCSEs did not mention modularisation, and instead concentrated on the change from coursework to controlled assessment (QCA, 2008). After all, by this stage A levels were modular in structure and the incoming Diplomas were modular. For the first time, almost all of the 2010 GCSEs contained four modules, which were available for assessment twice a year, with some subjects having modular assessments available more frequently. With many students taking 10 GCSEs, this meant 40 GCSE assessments plus the availability of one re-sit for each of the three non-terminal modules, giving a possible 70 assessments per student over the two-year course. Of course, not all students took every re-sitting opportunity (see the later section on re-sitting), but add to this the fact that students were entered for the examinations multiple times (see the later section on early and multiple entry). Prior to this, the qualifications were typically composed of two question papers, giving an average of approximately 20 examinations per student taking 10 GCSEs. Dominance of assessments over the student experience is apparent from these figures alone.

Modular examinations were not unanimously supported. Grade inflation was seen by the 2010 Coalition Government to have been caused, at least in part, by modular examinations, due to opportunities to re-sit and the bite-sized modular structure. Others blamed the accountability system and competition among examination boards for creating downward pressure on standards (Acquah, 2013; Cadwallader & Tremain, 2013; Heinrich & Stringer, 2012; House of Commons (HoC), 2012; Mansell, 2007; Meyer, 2011). Cadwallader and Tremain (2013) pointed to evidence (albeit limited) that accountability drove schools to approach assessment opportunities strategically. Modularisation facilitates new avenues for this, such as early entry, multiple entry to different examination boards' qualifications in the same subject and re-sits – strategies used particularly with low-performing students (Taylor, 2016).

Teachers had mixed views on modularisation, with some arguing that linear assessment promoted in-depth learning, especially related to subject specific skills, while others thought that students were better served by continuous assessment and being able to build on results (Alpha Plus, 2012). Re-sits and early entry were concerns for those who were uncertain about unitisation, given the performance target culture. Heinrich and Stringer (2012) found that the modular structure allowed schools to provide students on the grade C borderline with the extra feedback and tuition needed to increase their chances of gaining those critical C grades. The new Coalition Government set out the aspiration that the GCSEs would revert to linear-only routes and briefly considered the idea of replacing GCSEs in some subjects with English Baccalaureate Certificates to reinforce its proposition that GCSEs had been fatally 'dumbed down' (HoC, 2012). Changes to policy were announced before the new modular qualifications had a chance to be examined: modular examinations began in 2010; full qualifications were awarded in 2011. Following a DfE policy decision, Ofqual rendered examinations linear for GCSE courses awarded in summer 2014[10] and introduced new criteria for syllabuses in GCSE English language, English literature and mathematics for first teaching from September 2015. Reforms to other subjects followed from September 2016.

Despite the mixed views from stakeholders on the advantages and disadvantages of modularisation, GCSEs have been awarded in a linear fashion since summer 2014 and are now governed by different regulations, including a new 9 (highest) to 1 (lowest) grading scale. So, within this century, GCSE examinations have changed from largely linear, to all modular in structure, and reverted back to all linear in structure. Given the lack of evidence-based consideration of these policy changes, it is not surprising that some (e.g. Vidal Rodeiro & Nádas, 2011) have argued that the wholesale switch from linear to modular GCSE structure, was a "simplistic all or nothing approach" that had a "certain degree of trendiness" and was adopted without concrete proof that it improved standards.' One aim of the current research project is to investigate the contribution of modular GCSE examinations to the rise in national outcomes. As such, the effects of school accountability measures also need to be considered.

[10] http://comment.ofqual.gov.uk/gcse-reform/

School accountability and examination entry policies

School performance tables were introduced in England in 1992, as part of the then Prime Minister John Major's Citizen's Charter. Pupils' GCSE results are published to provide an indication of the outcomes of each school. These results are monitored by the government to identify failing schools, which might require intervention, or even closure. Over time, schools have also begun to use results to monitor teachers' performances and to include the results in performance-related pay systems. Meeting targets at teacher and school level have important consequences beyond the result for the child. All of these pressures affect behaviour in schools (Baird and Elliott, 2018).

Different statistical indicators have been set over the years. Initially, the objective was to improve the proportion of students attaining five good grades (C or above). This led to schools focusing upon students who might narrowly fail to gain a grade C at GCSE. However, many other qualifications (e.g. BTEC Tech Award, GNVQ) were credited as equivalent to GCSE examinations and counted towards schools' attainment targets. Thus, schools adapted their students' examination entry profiles to maximise their results and some subjects showed huge growth in their entries as a result. For example, GNVQ ICT[11] had the same number of points as four GCSEs. An attraction of modular examinations for schools was that students could re-sit assessments until they gained a result that took the school over the threshold of the target grade C.

Following a review of vocational qualifications by Alison Wolf (2011), the qualifications counted in the school performance tables were restricted. Governments have adapted the school performance measures to try to improve the impact of them upon schools' behaviours. As such, five GCSE results at grade C or above, including English and mathematics, became a target. Then from 2016, a new accountability measure was introduced: Progress 8.[12] This is a value-added measure, which takes into account students' prior attainment and produces a standardised score for progress in the mostly English Baccalaureate subjects although up to three other GCSE subjects or selected (sometimes modular) vocational qualifications can also be included. Now, schools also have two measures relating to the English Baccalaureate – one for the percentage of pupils entering and an average point score target for their performance.[13] Of course, GCSE examination outcomes are calibrated through the standard setting system, so it is to this that we now turn.

[11] Information and Communications Technology

[12] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/772801/Secondary_accountability_measures_guidance.pdf

[13] https://www.gov.uk/government/publications/english-baccalaureate-ebacc/english-baccalaureate-ebacc

Chapter 2 Standards

In large part, the impetus for the change in policy from modular to linear GCSE examination structures related to the perception that examination standards had been dumbed down. Here, we explain how standards are set and discuss a number of critical incident cases where modular examination structures played a part in public debacles.

Standard setting for GCSE examinations

GCSE examination standards are set separately for each syllabus. Until 2011, examination boards followed the regulator's *GCSE, GCE, Principal Learning and Project Code of Practice* (Ofqual, 2011). More recently the *Code* has been replaced with outcome-based rules, detailed guidance and for GCSEs a process of data exchange and monitoring. Both the guidance and the code list information sources to be taken into account in standard-setting meetings, including qualitative and quantitative evidence. This is conducted at 'awarding' meetings, which are presided over by the Chair of Examiners[14] for the syllabus and attended by other senior examiners and staff of the examination board. Most of the time in these meetings is spent with examiners scrutinising the quality of students' examination performances in relation to previous standards set. Many of these meetings are now held virtually.

Statistical information is also provided to the awarding meetings and is used to inform the standards set within the examination boards and then by the regulator in the data exchange and monitoring of outcomes. In the 1990s, GCSE standards were set with most of the weight in the decision making being placed upon examiners' qualitative judgments of students work. Research showed that these judgments were associated with biases and were imprecise (Baird, 2000; Baird, 2007; Baird & Scharaschkin, 2002; Cresswell, 1996; Forster, 2005; Scharaschkin & Baird, 2000). Ofqual introduced a 'comparable outcomes' methodology that provides statistical guidance for standard-setting from 2010 (Ofqual, 2014). The method predicts the cumulative distribution of grades at syllabus level based on candidates' prior attainment, using the relationship between prior attainment and performance at national level in that GCSE subject in previous years. Therefore, the comparable outcomes methodology assumes that progress in a particular subject is likely to remain stable over years at a population level so long as the entry does not change significantly.

As a methodology, the comparable outcomes technique is best suited to examinations in which a large component of the examination is taken at the end of the course. Standards cannot be controlled by grading at the end of the qualification if only a small proportion of the examination remains to be graded. If the pattern of entry for an examination is volatile, with students taking the examinations early and different proportions taking examinations at different stages in different years, the examination board has to monitor the effects of this to come to a decision regarding the utility of the comparable outcomes statistics.

[14] Examination boards now use different titles for senior examiner roles.

Whilst these principles apply regardless of the structure of the qualifications, the nature of the assessment in modular and linear qualifications has implications for the process by which standards are set and maintained. The following section discusses these issues, focusing on the challenges of setting and maintaining standards in modular qualifications.

Setting and maintaining standards in modular qualifications

As has previously been described, linear qualifications require students to sit all of the assessment at the end of the course. This means that each certificating student sits the same assessment at the same time, allowing grade boundaries to be set on the entirety of the assessment when students certificate. For modular qualifications, students are able (but not required) to sit modules of the qualification at different times throughout the duration of their course (depending upon the availability of the assessments). This means that certificating students will likely have taken modules at different times throughout the course, thereby having sat different question papers that require different grade boundaries.[15] Whilst ensuring overall qualification standards remains the priority in modular qualifications, it is also important that module standards are comparable such that students are not advantaged or disadvantaged depending upon when they take each module. This is challenging, and the complexities increase in line with the number of assessment opportunities that are available.

In modular assessment, grade boundaries for each module are typically set following each examination series, rather than on the entirety of the assessment when students certificate. This provides feedback to students on how they are performing throughout the course. However, it also places a burden on the standard setting process because grade boundaries must be set on each module of the qualification each time it is available that in combination should maintain overall qualification standards. Whilst much is known about the factors affecting the aggregation process, the empirical parameters vary with each set of modules aggregated, so it is not possible to entirely predict the effects of grading each module upon the overall qualification standard. Only when the results come together and the full data is known are the implications of the grading of each module fully available.

In England, the process for setting standards on the modules of a modular qualification generally follow a similar approach to setting overall qualification level standards: statistical predictions (the comparable outcomes approach (Taylor and Opposs, 2018)) are generated that, in conjunction with senior examiners' qualitative judgment of students' work, guide the setting of grade boundaries on each module. The first challenge of maintaining standards in modular qualifications lies in generating these statistical predictions.

[15] This also requires a mechanism for combining marks that students have achieved in different examination systems. For GCSEs and A levels this has been achieved by the uniform mark scale (UMS) that standardises students' marks before combining them.

Generating statistical predictions to guide standard setting at the overall qualification level is generally fairly straightforward in the sense that the predictions model the relationship between prior attainment and outcomes for a common group of students who are certificating at the end of their course (at GCSE, this is typically 16-year-olds). In modular qualifications, the situation is likely to be different: students can sit modules at different times, meaning that a GCSE student might sit the first module of a qualification up to 18 months before they certificate. This raises issues for generating statistical predictions at module level, particularly if there are no existing module standards on which to base these predictions (such as in the first year of a new qualification). Here, the predictions must be based on the final subject outcomes instead.

Characteristics of the students

The first issue in generating statistical predictions at module level therefore stems from the likely differences in the characteristics of the students who are sitting an early module of the qualification and students certificating at the end of the course. In a modular qualification, students are likely to sit some modules of the qualification prior to the series in which they certificate – indeed, the very nature of modular qualifications encourages this. Thus, when generating module predictions based on subject outcomes there would likely be differences in the characteristics of the students for whom the predictions are being generated and on whom they are based. For example, at the extreme, one might be trying to predict the performance of a 15-year-old student who has had three months of teaching, based on the outcomes of a 16-year-old student who has had two years teaching. Predictions are based upon prior attainment, but it is clear that even where such students have the same prior attainment, performance is likely to differ.

Differences in performance are likely to result from a number of factors. First, students entering at different time points will have had different amounts of teaching and exposure to the subject content, which will likely impact on performance. Second, students are likely to mature during a two-year course of study and therefore perform differently depending upon when they take the assessment (Clark, 1996; Taverner & Wright, 1997; Vidal Rodeiro & Nádas, 2012). Finally, students sitting the module at different times are likely to have different levels of motivation: there is some evidence to suggest that students tend to take an early assessment attempt less seriously if they know that there is an opportunity to re-sit (Heinrich & Stringer, 2012; see later discussion). Furthermore, some schools might use the early modules as 'mock' examinations to assess how their students are progressing. The extent to which these factors might influence performance is difficult to quantify, meaning that generating appropriate statistical predictions based upon prior attainment at the module level is challenging.

Aggregation effects

A second issue in generating module level predictions relates to the aggregation effects that occur when module marks are combined to qualification level. In any

qualification with multiple assessments students typically perform differently across each element of the assessment: a high-performing student is unlikely to perform consistently well on all of the assessments, and a low-performing student is unlikely to perform consistently poorly on all of the assessments. This means that when aggregating multiple assessments, 'regression to the mean' can occur, where there is a tendency for students to regress towards the middle of the mark distribution (the same phenomenon occurs in repeated testing; see Smith & Smith, 2005). This means that to achieve a desired outcome at qualification level – for example for 20 per cent of students to achieve a grade A – the module outcomes typically need to exceed this. Due to the effects of regression to the mean, the opposite effect occurs at the lower end of the grade distribution – for example at grade F, module outcomes typically need to be lower than the desired qualification level outcome. The effects of regression to the mean differ depending on the number of modules in a qualification. As such, individual examination boards are likely to need to set different module level standards in the same subjects to account for the structure of their qualification.

Generating module level predictions based on subject level outcomes does not, or indeed cannot, account for this effect because the extent to which it will occur for a particular set of data is unknown. Correlations between module outcomes vary. This means that if module predictions are not adjusted, there will likely be too few students achieving the top grades and too many achieving the lower grades. This can result in the standards of the final (or terminal) assessment having the burden of ensuring that overall qualification standards are maintained. To mitigate these risks, one approach is to try to model the likely effects of regression to the mean, then adjust the module standards accordingly. However, any modelling is imprecise so the risk remains.

Of course 'regression to the mean' also exists in a linear context where students take multiple assessments. However, the difference is that the effects on overall outcomes are known in the linear structure and can be accounted for when setting grade boundaries, since students sit all of the assessments concurrently. Often modular structures have more components too and the larger the number of components, the greater the regression to the mean effect.

Re-sitting effects in standard setting

Below we discuss the research on re-sitting more fully. Here we consider the effects of re-sitting on standard setting processes specifically. In modular qualifications, there is generally a facility to re-sit some (or all) of the modules of the qualification, with research suggesting that students typically improve their performance on the re-sit (Vidal Rodeiro & Nádas, 2012). Whatever the reason for a change in performance, the facility to re-sit means that even if standards have been set appropriately on each module initially, outcomes are likely to shift as some students choose to re-sit the assessments and perform differently (note that this can only be in an upwards direction, since students receive their best mark as their final mark). Similar to estimating any effects of maturation or regression to the mean, quantifying the effects of re-sitting or taking this into account when setting grade boundaries is difficult: it is not known how many students will choose to re-sit, which students will re-sit (see Vidal Rodeiro &

Nádas, 2012), when they will re-sit, or how they will perform. In the modular GCSEs introduced in 2009, students were only able to re-sit each module once throughout the duration of their course. Whilst this approach affords some control over the extent of any re-sitting effect, the very presence of re-sitting creates challenges for maintaining standards in modular qualifications.[16]

Low weighting of final assessment – banked marks

A further issue in setting grade boundaries in modular assessments relates to that of 'banked' marks. As grade boundaries are generally set each time a module is available, results are provided to schools and students at the same time. These standardised marks essentially become 'banked' in the sense that they cannot change (unless students re-sit, when their mark could increase). The facility to bank marks means that students could certificate in a given series without having to sit any assessment – they would simply be 'cashing-in' all of their previous modules. For these students, it would not be possible to influence their final outcome since their grades have already been awarded. This can be problematic and can compromise overall qualification standards, as in the case of GCSE science in 2008 (see Box A at the end of this chapter).

New GCSE science specifications were developed for first teaching in 2006 which replaced double award science with a number of alternative qualifications: science, additional science, additional applied science and the single sciences (biology, chemistry and physics). The regulatory criteria for these specifications required them to be modular in structure, but the proportion of internal and external assessment was flexible (Ofqual, 2009a). Examination boards interpreted the criteria differently (as was allowed) and developed qualifications with differing numbers of external examinations – from two to six. Furthermore, these qualifications had different weightings for internal and external assessment and different opportunities for re-sitting modules.

The high degree of optionality and routes through the qualifications led to difficulties in ensuring that overall standards were maintained when these qualifications were first awarded in 2008. In some examination boards, students were able to re-sit modules an unlimited number of times and bank a large proportion of marks prior to the series in which they certificated. This meant that the ability to influence the final grade these students achieved was limited, leading to difficulties in maintaining standards. Whilst these difficulties were not experienced by all examination boards, they ultimately led to one of the other examination boards being requested to lower its grade boundaries to ensure that standards were aligned, thereby diluting the overall GCSE science standard. This prompted the examination regulator to require examination boards to implement a number of changes for the following year (Ofqual, 2009b), though concerns still remained (Ofqual, 2010). Ultimately, new GCSE science specifications were developed for first teaching in 2009 against tighter criteria, including rules around re-sitting and the amount of assessment that students could sit prior to the series in which they certificated – generally known as the 'terminal rule' (see Appendix B).

[16] This is not to suggest that re-sitting should not be allowed, rather it is to demonstrate the challenges it creates.

The introduction of a 'terminal rule' is typically used to mitigate the risk of 'banked' marks compromising standards, and was put in place for all of the modular GCSEs first taught in 2009. In this case, the terminal rule required students to sit at least 40 per cent of the overall assessment in the series that they certificated in. Whilst this mitigated the risk of students banking marks for all of the assessment prior to the series that they certificated in, there were still issues. The terminal rule specified the amount of assessment that students must sit in the series that they certificate, but it did not dictate which 40 per cent of the assessment this should be. Therefore, depending on the structure of the qualifications, it might be possible for students to use different modules as their terminal assessment. This can introduce complexities for maintaining standards since changing the grade boundaries on different modules will impact differently on students, depending on which unit students are using as their terminal assessment – this will influence the final rank ordering of students.

The terminal rule can be particularly problematic when it is possible to use different forms of assessment – in particular, internal assessment – as the terminal module. The practice of setting grade boundaries on each module at the time that students sit the assessment means that schools and students are fully aware of how students are performing, even to the extent that, for some qualifications, they are able to calculate the number of marks that a student needs to achieve in their final module to achieve a certain grade overall. This can be problematic when the final assessment could be coursework or controlled assessment, as in the case of GCSE English and GCSE English Language in 2012. Although a number of factors were at play, the modularisation and the structure of the assessment played a significant role in the events that unfolded.

One approach to mitigate the risks of setting grade boundaries on early modules is to allow students to sit modules early (thereby preserving the modular approach to assessment), but to withhold results or not set any grade boundaries until the series that students certificate. This approach was used in GCSE English/English language in 2013, following the events of 2012: students were able to enter the assessments in the January series and they were marked, but they were not graded until June. This approach facilitates the setting of grade boundaries on each module when students certificate since there are no banked marks. It also mitigates some of the risks of having to generate statistical predictions for the early modules. However, this approach is not likely to be universally welcomed. Schools and colleges are likely to use the performance of students on the early modules as feedback – to decide whether students should re-sit a module in a later examination series and to identify areas where students might need further support to improve their performance. Without this feedback, such decisions would be more difficult.

Alternative syllabuses

The discussion above has highlighted some of the key challenges in setting and maintaining standards in modular qualifications. This situation can be further complicated where multiple qualifications with alternative routes to certification exist concurrently. For example, where both modular and linear versions of a

qualification are available, it is possible for students to sit some of the modules of the modular qualification, then, depending on their performance, choose not to certificate in this qualification, instead favouring the linear qualification where all the assessment is at the end of the course. This approach can be beneficial to students since it essentially 'wipes the slate clean' and allows students to sit the whole assessment without carrying forward any of their marks from the early modules. It is therefore likely to be favoured by students who under-performed on the early modules and were not on course to achieve their target grade (Taylor, 2016).

A consequence of this, however, is that the subset of students who certificate in the modular qualification are unlikely to be representative of those who sat the early modules (and those on whom the standards for the early modules were set). This means that whilst appropriate standards may have been set on the early modules for the whole cohort of students entering them, these students are unlikely to be representative of those that choose to certificate. As such, inappropriate module standards may have been set when only the students that go on to certificate are considered. This can create complexities when trying to meet statistical predictions at the qualification level and may result in the final (or terminal) assessment standards being distorted.

Examiner judgments

Standard setting procedures involve not only statistical information, but qualitative judgments of the standards of students' work by subject matter expert examiners. Ultimately, the goal is to maintain the qualification standards when examiner judgments are made on cut-scores for individual assessments. The effects of aggregation can result in different proportions of students gaining the grades than the examiners intended and this is exacerbated by modular qualification standard setting. For linear examinations, the aggregation effects can be modelled in the meetings and the results can influence the decisions. However, for modular qualifications decisions have to be taken part-way through the qualification when their effects upon qualification standards are to some extent unknown. Fluctuating entry patterns make the effects of aggregation even less predictable. In early years of modular qualifications, teachers' entry policies were variable and some candidates were not adequately prepared for the examinations. Thus, examiners had to form judgments about the difficulty of the examination papers versus the preparation of candidates without knowing how these candidates would perform on subsequent assessments and how this would affect their overall grades.

Case studies in grading modular GCSE examinations

Standards for GCSE examinations are set and defined at the level of the syllabus. If students have a lot of module results 'in the bank', but do not certificate (also known as 'cashing in'), then the effect of those modular results upon overall qualification standards are essentially unknown and can be difficult to model or predict. These issues caused severe problems with the introduction of the

Curriculum 2000 A level, which led to a Public Inquiry, and were part of the problem for the GCSE English examination in 2012 which led to a Judicial Review (see Appendix B). Where problems arise in the introduction of examinations, they tend to have multiple causes. So it was with the introduction of the new GCSE English examinations in 2012. We outline the issues in some detail below to illustrate how changes to the amount of coursework ('controlled assessment'), together with modular assessment and various other factors interacted to produce the problems encountered when the examination results were certificated.

After many years of a steadily increasing proportion of students achieving the all important grade C and above in GCSE English, in the summer of 2012 the overall proportion decreased, with a reduction of 1.5 percentage points of students gaining at least a grade C compared with the year before. The difference in headline results was explained by a change in the ability profile of the cohort entered. Importantly, the drop in results was not uniform – some schools saw significant increases, others significant decreases in their students' outcomes (Ofqual, 2012) with sometimes potentially serious consequences for the latter due to accountability considerations. This led to a public outcry. Although the new examinations were modular, factors interacted to produce the outcomes. This case shows the complexity of these interactions.

Controlled assessment is similar to coursework in that the tasks remain similar (or the same) between each examination series, often meaning that grade boundaries also remain the same. Thus, there was an expectation from schools that the grade boundaries for the GCSE English and GCSE English language controlled assessment in June 2012 would remain the same as those set in previous series. Based on this assumption (and for those students using the controlled assessment as the terminal assessment), schools were therefore able to calculate the number of marks students required on the controlled assessment to achieve a certain grade – often the threshold grade C. This knowledge is likely to have influenced the marking and administration of the controlled assessment units within schools – assessments that are internally marked. Indeed, analyses suggest that of the students using the controlled assessment as their terminal assessment, a large proportion would have achieved a grade C using the controlled assessment boundaries from previous series, but fell below this threshold when the controlled assessment boundaries were increased in June 2012 (AQA, 2012a; Ofqual, 2012).

The changes to the controlled assessment boundaries resulted in several schools receiving unexpected results: some schools had reportedly informed students what their final grade would be prior to results being issued, based on the assumption that the controlled assessment boundaries would remain the same. The public outcry that followed culminated in a Judicial Review of the grading process brought by a consortium of schools and local authorities. This ultimately found in favour of the examination boards and Ofqual, citing the complexity of the qualification as key factors in the grading issues. A number of changes were made to mitigate these risks in future series, yet this case clearly demonstrates the risks associated with maintaining standards in modular qualifications.

Setting standards for GCSE English in 2012 was therefore a complex task in which the modularisation of GCSEs played a key part. GCSE science standard setting had previously proved to be difficult (Box A), for reasons entirely related to the unitised structure, but in this case it was because of the 'external' approach to the modular design (see Table 1.1), which entails modules being available across qualifications.

Box A The trouble with choice: modular science GCSEs

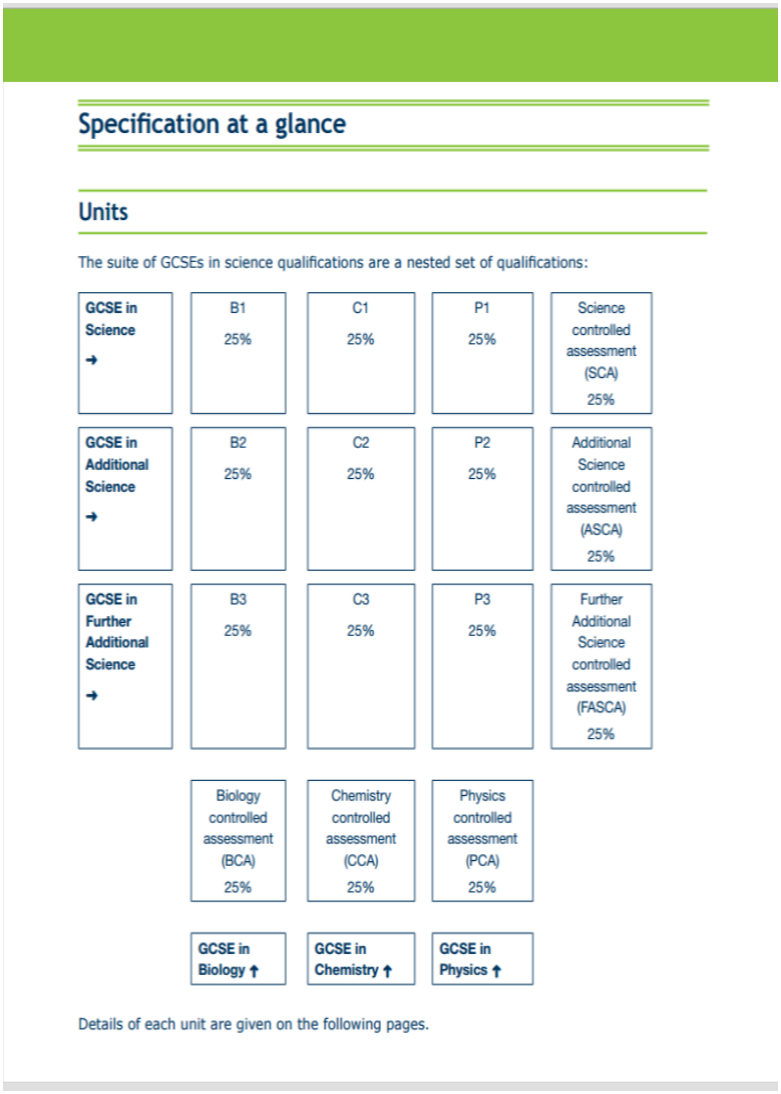
In 2006, following a brief pilot based on University of York recommendations about the need to include in science specifications both traditional science for those who might progress to study A level science, as well as science for societal participation ('how science works'), new GCSE science specifications were first taught that included single sciences (biology, chemistry and physics), science, additional science and additional applied science (the latter three in place of double award science). The regulatory criteria were flexible about proportions of internal and external assessment and also, while specifying that qualifications should be 'unit based', left open to examination boards how many units there should be (Ofqual, 2009a). Examination boards interpreted the criteria differently, as was allowed, and offered differing numbers of examinations – from two to six – as well as different weightings for internal and external assessment. In one examination board, students were also allowed a choice of assessment routes. Initially, students were able to re-sit units throughout their two-year course an unlimited number of times.

Difficulties in maintaining standards were experienced by some examination boards, which led to one of the examination boards being required to lower its grade boundaries to bring its standards in line with that of the other boards. Ofqual (2009b) reviewed the comparability of the different specifications because of the different forms of assessment, different routes students could take and the possibility that the highest achieving students might not be stretched due to the low weightings of external assessments (in cases where there were a plethora of papers). Ofqual required that examination boards introduce more demanding assessments and rescinded much of the flexibility that unitised science allowed when it introduced regulations for the full suite of GCSE subjects. GCSE specifications that were accredited for first teaching in 2009 and 2010 all contained four modules, had terminal assessment and re-sit rules. Science GCSE assessments were re-developed along the lines of the other GCSEs that were introduced for first teaching in September 2009 and new criteria were published in 2011.

Most of the discussion around science GCSEs related to the balance of internal and external assessment and the inclusion of 21st century science and how science works, the economic and strategic role of science, the role of vocational science, the role of practicals and progression to A level sciences (see Black et al, 2004; Donnelly & Ryder, 2011; Horner & Ryder, 2014; Millar, 2011, 2012; Ryder & Banner, 2011, 2013). Ryder and Banner's (2013) survey of teachers' reactions to the 2006 changes did not even mention unitised assessment methods and none of the teachers interviewed offered an opinion on them.

Although some examination boards (for example Pearson) had identical examinations for both three qualification awards (i.e. biology, chemistry and physics and science, additional science and further additional science) others (for example AQA) had slightly different examinations (80% crossover between separate subjects). Pearson's diagram of the different routes is below (Figure 2.1).

Figure 2.1 Routes through GCSE science (Pearson 2011)



Specification at a glance

Units

The suite of GCSEs in science qualifications are a nested set of qualifications:

GCSE in Science →	B1 25%	C1 25%	P1 25%	Science controlled assessment (SCA) 25%
GCSE in Additional Science →	B2 25%	C2 25%	P2 25%	Additional Science controlled assessment (ASCA) 25%
GCSE in Further Additional Science →	B3 25%	C3 25%	P3 25%	Further Additional Science controlled assessment (FASCA) 25%
	Biology controlled assessment (BCA) 25%	Chemistry controlled assessment (CCA) 25%	Physics controlled assessment (PCA) 25%	
	GCSE in Biology ↑	**GCSE in Chemistry** ↑	**GCSE in Physics** ↑	

Details of each unit are given on the following pages.

As previously mentioned, GCSEs were not the first modular examinations. In the next section, we turn to modular assessment in higher education, vocational education and in other national examinations.

Chapter 3 Modular assessment in other settings

Higher education

Higher education (HE) is now fundamentally modular. Its introduction in this setting predates its use in national qualifications. HE assessment meets the definition of modularisation and assessment as 'the idea that the curriculum can, and even should, be broken down into more discrete units of accessible study' (Smith & Bradley, 1996) has been widely accepted. Even the most linear degrees, for example the Cambridge Tripos, are divided into separate stages based on the three years of study. This is a looser, and familiar, form of modularisation. As Bekhradnia (2004, p.5) observed, 'Modularization ought not therefore to be a controversial notion, where controversy does arise is around the length and intensity of modules'. Tight frameworks have produced more debate, as discussed below. In the UK, one pioneer of credit-based modular systems has been the Open University that, since its inauguration in 1971, has operated a system of modular credits which are accumulated into a final degree.

In the United States, credit accumulation and transfer (CAT) frameworks have a longer pedigree with their origins dating back to the late nineteenth century, when there were moves towards a more flexible university curriculum:

> At a philosophical level, there was a growing acceptance of student-centred learning and of John Dewey's advocacy of self-realisation achieved through study fitting the individual's interest. There was also increasing demand for courses of a practical nature relevant to the real world'
>
> (Theodossin, 1986, p.5)

In this context the development of modular credit systems in the United States was a natural outcome.

In the UK, development of CAT frameworks has been more contested. The 1992 ending of the binary divide in which colleges and polytechnics were incorporated into the university system led to considerable debate about the value of modularity. This was because modules were integral to the degrees offered in the 'new universities' but unfamiliar in most traditional universities. Betts and Smith (1998), writing at a time when over 90 per cent of universities had adopted a modular or unitised curriculum recognised 'that the UK credit-based modular systems are based on a philosophy that has its roots only a few inches below the surface. It runs counter to HE philosophy and ideology that has withstood the test of time' (p.5). They argued that the introduction of credit-based modular systems (CBMS), which originated in the colleges and polytechnics, were resisted because of the challenges they brought:

> The dubious, inconsistent and arbitrary practices of the universities' examinations boards are now exposed through the transparency of the CBMS assessment system. The difficulty of establishing comparability between subject disciplines and

> institutions, the weaknesses of the external examiner system and the precarious nature of the classification system have all been present for many years. (p.5)

The advocacy of the credit-based modular approach was based in a confidence in the clarity of the specifications and assessment requirements that were a necessary part of course recognition by bodies such as the Council for National Academic Awards (CNAA) who were empowered to grant institutions their own degree awarding powers. The CNAA sought to maintain comparability across sectors and between universities by requiring detailed specifications and assessment information.

Atkins et al. (1993) summarised criticisms of this position as follows:

> uncontrolled modularization will weaken the quality of learning achieved in higher education. For example, although some subjects can be studied on a 'cafeteria' basis, others require a linear progression through a hierarchy of concepts of increasing cognitive complexity and difficulty. The growth of general cognitive skills and personal competences may also be better served through a developmental rather than an accumulative framework. Further, in modular schemes, summative assessments are likely to occur more frequently while the opportunities to remediate learning deficiencies may become less.

Brown and Saunders (1995), themselves sympathetic to increased modularisation of courses, cited similar issues faced by institutions which sought to move in this direction. They found the main criticisms from academic staff were:

> (1) the *management* demands, and confusions, around operating such a system – particularly if they were accompanied by semesterisation (the shift from three terms to two teaching blocks per year);
> (2) the impact on *choice* – with the impact of timetabling and professional body requirements often reducing choice;
> (3) the risk of *academic fragmentation* – in which students did not synthesise their learning across the modules;
> (4) the increase in *assessment* which increased demands on student time and reduced teaching time.

Given these reservations about the impact of modularisation, why has it been so widely adopted? The answer can largely be seen in terms of wider participation, flexibility and transferability. The central driver has been that of widening participation which has increasingly involved non-traditional and part-time students who may not be in a position to follow a three-year linear programme. Historically this was recognised in both the Robbins Report (1963) and the Dearing Report (1997), with Dearing concluding that:

> If the higher education sector is to be truly committed to the concept of lifelong learning, students will need to take advantage of a national system of credit accumulation and transfer. Transfer will not be possible without some level of national currency of the credit acquired by the student.
>
> (Dearing, 1997, 10.62)

This recognised the changing nature of the HE student population which had increased dramatically and by 1994 included 50 per cent 'non-traditional' students who were over 25 and largely part time (Slowey, 1995).

The move from elitist to mass HE has also required more flexibility in the system. This may mean a student studying part time and needing to accumulate credits over years, sometimes with interruptions. Part of this is transferability across systems. In California over 50 per cent of students in the state university system will have spent two years in a local junior or community college and will carry over credits from their programmes of study which exempts them from the first year of their university undergraduate course.

One of the most ambitious transfer schemes is that of the EU's Bologna Process which since 1999 has been encouraging a greater degree of convergence in HE across the 46 countries participating (Adelman, 2008). The aim has been to harmonise courses, credentials and qualification frameworks so that they will be recognised across borders and allow significant international mobility. One manifestation of this is the ERASMUS programme (now Erasmus Plus), which allows students to study in other EU countries. In 2012-13 alone, 270,000 students took part.

The arguments for and against stronger forms of modularisation, for example credit-based modular systems, show some commonality with those in general and vocational education (see below). There is common support for more explicit and manageable units of study and recognition of the need for more flexibility given the changing nature of the student population. This is more pronounced in HE where the increase in non-traditional students has been dramatic. The criticisms are also similar: the encouragement of atomised learning which does not lead to a grasp of the subject as a whole, the risk of over-assessment and the reduction in teaching time available. Further, grade inflation in HE has been linked with modular assessment as a possible cause (Bachan, 2015). As well as HE, modular assessment has been a feature of vocational education for many years and we next outline the developments in this area in the UK.

Vocational education and training

Since the introduction of school education for all in the UK, there has been an ongoing tension about the suitability of academic qualifications for the entire cohort. Vocational education at school level has often been portrayed as a solution for positive achievement with utility for less academic children (Wolf, 2002). Arguments for the unified frameworks discussed above pertain in this

sector too (Howieson, Raffe, Spours, & Young, 1997). Standardised and flexible approaches to education and training are needed for the knowledge economy, it is argued (Ertl & Hayward, 2010).

GNVQs, NVQs[17] and their Scottish equivalents were all developed in modular forms (Young, 1995). They have had a chequered history of frequent syllabus and examination revisions due to controversies over standards and their equivalences with more academic qualifications. The GNVQ, which was first developed in the early 1990s and was available until the early 2000s, particularly suffered this fate. Originally conceived as a post-16 qualification only, it soon became available in truncated form (three units instead of the full six for the GCSE equivalent foundation and intermediate levels) for Key Stage 4 students. Issues around 'parity of esteem' with A levels and GCSEs dogged GNVQs from the outset; its initial teacher-marked, portfolio-based, specific outcomes driven, ungraded mastery assessment model quickly morphed into something that more resembled its academic cousins. Achievement on each module at first relied on the successful completion of highly specific performance criteria (outcomes) plus passing at 70 per cent or higher a fairly straightforward multiple choice test. By the late 1990s the tests were abandoned and the performance criteria were replaced with GCSE and A level-style assessment objectives. The GNVQs themselves were replaced in 2003 with applied GCSEs and A levels, which, while remaining unitised, were graded A*-G and A- E respectively.

GNVQ achievement was reported in performance tables starting in 1996, but at that point the results were reported separately. The then Labour Government strongly believed in the efficacy of vocationally-related qualifications and a raft of these qualifications were given performance table points – some of them equivalent to four GCSEs in 2003 (Isaacs, 2013a). Many schools expanded their vocational offerings because of the high equivalences some vocational qualifications received, and the number of Key Stage 4 students taking such qualifications expanded from 1,882 in 2003/4 to 462,182 in 2009/10 (Wolf, 2011). The *Wolf Report* (Wolf, 2011) decried the value of many vocational qualifications for young people's economic benefit and many of them were reweighted or removed from the performance tables by the then coalition Government.

The qualification developed to replace GNVQs and other national applied qualifications (and, it was hoped, other awarding organisations' own brand vocational suites) was the short-lived Diploma qualification. It, too, was modular in nature. Introduced in 2005 under the Labour Government, these two-year qualifications were first examined in 2009, with initial hopes for large candidate numbers proving largely ephemeral. Developers of Diploma policy and construction never learned the lessons from GNVQ and the qualification's complex structure – where candidates had to pick and choose from a long list of units and qualifications in addition to the mandatory units – and hurdles-based assessment model, meant that the seeds of its own failure were built in to the

[17] National Vocational Qualification, a work-based qualification which recognises the skills and knowledge a person needs to do a particular job.

very core of the model. Once the coalition Government came to power in 2010, the Diplomas were dropped (Isaacs, 2013b).

While applied GCSEs in their modular format remained available for a few years after 2010, when the reform of the entire GCSE suite was announced GCSEs in applied business, applied ICT, health and social care, hospitality and catering, leisure and tourism, manufacturing and performing arts were dropped. This ended *centrally-devised* modular vocational qualifications for Key Stage 4 students, although awarding organisations continue to offer own-brand modular qualifications at levels 1 and 2 (Ofqual, 2015b).

End of course, examination-style assessments do not always fit well with vocational syllabus aims, which are often better aligned with authentic assessments and testing when ready approaches. For example, vocational assessments frequently include practical tasks (see Richardson & Sing, 2011). Competency-based, outcomes-based statements of standards, in which all of the students could pass if they reached the required standard, are also at odds with the selective purposes of academic examinations. Further, outcomes-based approaches have been criticised as being too narrow to be properly educational because they sometimes focus overly upon discrete skills or vocational knowledge that is not generalizable (Wolf, 1995; Young, 1995).

In summary, modular assessment formats are the norm in higher and vocational education. Our next question is whether they are to be found in national assessments in other countries, or whether England is distinctive.

Use of modular examination structures in national assessments

In 2017, we researched summative assessment arrangements at lower and upper secondary (including university entrance examinations) in 18 jurisdictions, plus the International Baccalaureate. The predominant pattern for summative assessment at the end of a phase (lower or upper secondary) is a set of linear examinations rather than assessment in modular fashion building up to an overall outcome (see Appendix C). Queensland's upper secondary, France's lower secondary and Scotland's lower and upper secondary assessments were the exceptions. However, we must point out a major caveat. For many of these systems, teacher judgment has a direct influence on students' overall grades, at the very least through coursework, but in some cases, as in Ontario and the US states, determining the grades themselves. Thus, there might be a form of staged assessment in operation in those systems. Box B outlines the use of modular examinations in Scotland's public examination system since it, too, was unusual in its modular nature until recently.[18] Indeed, assessment reform is so common across the globe that descriptions of examination systems often include ongoing reforms. In 2018, the National 5 assessments were made linear, the same process is occurring for the Higher examinations in 2019 and the Advanced Higher examinations will be made linear next year. The National 4 assessments

[18] https://www.sqa.org.uk/sqa/78398.7831.html

remain modular for now. Following this, we turn to the introduction of modular A levels in England.

Box B Scotland's examination system (2017)

Scotland's assessment system has its roots in *Opportunity for All* (1994) and the Higher Still programme. The programme merged academic and vocational courses with national qualifications at Access, Intermediate, Higher and Advanced Higher levels (McVittie, 2008). It included stand alone internally assessed units that were either qualifications in their own right or were combined in National Courses. These were graded pass/fail, but to achieve a full award, students had to pass all units plus an external assessment. Internal assessments could be re-assessed one or sometimes two times. Some teachers saw both internal assessment and the opportunity to re-visit work as an unwarranted increase in their workload (McVittie, 2008).

Modularisation was adapted from post-compulsory vocational education and was seen as a means of encouraging flexibility, credit accumulation and transfer and identification of common elements in different courses. Policy dialogue centred on risks related to fragmentation, over-assessment and assessment-driven learning. The Higher Still programme consisted of units that could stand alone or be combined into courses, although the units at Standard Grade were not certificated separately. These units were incorporated into National Courses in which internally assessed units were externally verified and were complemented by external examinations. Unit performance was ungraded – the overall course grade was based on the examination grade.

The *Curriculum for Excellence* superseded the Higher Still programme in 2010 although much of the structure of lower and upper secondary assessment remains in place. National Courses are available at levels 1 to 5, Higher and Advanced Higher. To get an award for a course above National 1, candidates must pass all units, which are internally assessed. National 5, Higher and Advanced Higher courses require unit achievement (internally assessed) plus a graded external assessment, which can consist of an examination, performance, project, practical activity or other form of assessment, and are graded A-D. The rationale behind the assessment regime is increased teacher flexibility and best use of professional judgment to ensure students have reached appropriate standards. Naturally occurring evidence can be used, which is supposed to result in less assessment. National 4 courses include an Added Value Unit (internally assessed) and National 5, Higher and Advanced Higher Courses include a Course Assessment that is externally marked by the Scottish Qualifications Authority (SQA) or internally assessed. Most National 5, Higher and Advanced Higher courses are assessed by a combination of examinations and coursework. There are no re-assessment rules as such, but teachers are encouraged not to give students more than one or in exceptional circumstances two chances to resubmit. Assessment methods include assignments, case studies, practical activities, performance, portfolio, project and examinations. Most coursework and National 5 and Higher assessments are set and marked by SQA; at Advanced Higher coursework is more likely to be set and marked as a shared responsibility between SQA and schools (SQA, 2013). Critiques of *Curriculum for Excellence* have focused upon curriculum, not its modular nature (e.g. Priestley et al., 2013).

GCE Advanced Levels

Modularity in GCSE in 2007 was anticipated many years earlier through the incremental, and then wholesale modularisation of GCE A levels. Modular A levels, especially in mathematics and science, predated Curriculum 2000 by over a decade (Hayward & McNicholl, 2007; Holding, 1995; Taverner & Wright, 1997; Young & Leney, 1997). Regulatory control by the School Curriculum and Assessment Authority (SCAA) of modular A levels began in 1993, introducing the notion of Advanced Supplementary (AS) alongside full A levels. This AS was half as large as the A level and assessed at the same (end of two years) standard (Holding, 1995). The first group of subjects was taught from September 1994 and featured synoptic assessment, maximum of 20 per cent coursework (some of the earliest modular syllabuses had 50 per cent coursework (Young & Leney, 1997)), 30 per cent terminal assessment and the ability to re-sit. By 1997 all but two of the eight examination boards offered modular A levels in mathematics (Taverner & Wright, 1997). Mathematics was a particularly attractive subject for modularisation because students could have more choice of the type of mathematics they wanted to study beyond core (pure) mathematics.

However, a big regulatory change took effect in light of the Dearing Review (1996), written for John Major's Conservative Government and largely adopted by Tony Blair's Labour Government through Qualifying for Success (QCA, 1998). The national framework of qualifications (NQF) was introduced as the organising principle through which all qualifications could express, at minimum, standards (level) and size (notional teaching and learning hours). For A levels, this meant the introduction of an Advanced Subsidiary (AS) qualification that, like its predecessor was half the size of a full A level, but this time assessed at the end of first year standard and crucially aligned with the Advanced General National Vocational Qualification (AGNVQ), which was a six unit qualification. AS qualifications contained three modules; full A levels, six. Behind these changes were the notions that students would broaden their studies, starting out on four subjects rather than three, and be able to mix and match vocational and general qualifications. The Labour Government was keen to promote 'parity of esteem' for vocational study (Priestley, 2003; Hodgson, Spours & Waring, 2005; Hayward & McNicholl, 2007). It was thought that students might take more science, technology, engineering and mathematics (STEM) subjects if they knew they were able to cash-in an AS qualification if things did not go as planned, and that teachers would be able to use AS results formatively in preparation for second year, A2 study.

The immediate outcomes of the introduction of Curriculum 2000 have been thoroughly discussed elsewhere (Alpha Plus, 2012; Hodgson & Spours, 2005; Hodgson, Spours & Waring, 2005). In 2001 timetabling problems arose and students (and teachers) complained vociferously about assessment overload (students were at the time expected to take four AS levels), the loss of weeks of teaching time at the end of Year 12 and a decrease in students' extra-curricular activities. Few students were mixing and matching qualifications, which had been a selling point for the original scheme. Students' results rose in 2002, but that outcome was seen in a negative (considered too easy) as well as a positive

light. New A levels introduced for first teaching in 2008 tackled some of the problems. For the most part the number of modules shrank from six to four (although not in mathematics and the sciences), stretch and challenge were introduced in the A2 units with the intention of making them more demanding, and a new A* grade was introduced based mostly on the outcomes of the A2 units. These reforms were intended to reinforce both breadth and depth of study. January module examinations were withdrawn by examination boards in cases where there had been low numbers of entrants, which essentially restricted re-sitting. In 2014, the January examination series was removed. Regardless of these changes, the Coalition Government announced the reintroduction of linear assessment in A levels, with the new syllabuses in 'facilitating' subjects[19] (aside from mathematics) being available for first teaching in September 2015.

Schools Minister David Laws explained the rationale for the changes, as follows:

> ...[W]e want to give students a better experience of post-16 study, ensuring they are studying for rigorous qualifications that will provide them with the right skills and knowledge to allow them to progress. Students currently start A levels in September and then they immediately start preparing for examinations in January. They and their teachers have spent too much time thinking about exams and re-sitting them, encouraging in some cases a 'learn and forget' approach .... The old rules allowed multiple re-sitting of those papers, so a student might sit some papers in January, and if they wanted to improve their grades they could re-sit them in June and again the following year, while sitting and then re-sitting their A2 papers. In 2010, 74% of maths A level students re-sat at least one paper.
>
> During the past few years, too many students in our schools system have spent too long preparing for and taking tests in years 10, 11, 12 and 13. During the past decade, we have been in danger of creating an "exam factory" in our schools, particularly in the last four years of education, rather than creating places of deep learning where teachers and students are given the time and space to develop deep knowledge of subjects, rather than just preparing constantly for public examinations.
>
> The focus that there has been on exams in every one of those final four years of school education can lead to young people failing to deliver and develop that deep understanding of their subject, and to their failing to make connections between topics. Re-sits have also led to too much teaching time being sacrificed for assessment preparation.
>
> (quoted in Long, 2015)

[19] http://russellgroup.ac.uk/for-students/school-and-college-in-the-uk/subject-choices-at-school-and-college/

Benefits and drawbacks of the modular A level system were rehearsed from the late 1980s (Moon, 1988; Warwick, 1988; Watkins, 1987) with no clear consensus, and continue to be debated. Early on, Holding (1995) believed that the modular system motivated and encouraged students to study hard from the outset of their courses, with colleges able to run them efficiently and better retain post 16 students. He saw its main drawback as the assessment level of the AS because it was to be assessed at the full A level standard. As students were only part way through the course, this could be difficult for them to attain. In the event, this was not problematical. Hirst and Meacock (1999) studied the mathematics intake for Southampton University noting that while increasing numbers of students were entering from a modular route, there were no statistically significant differences between them and the group who had done the linear route, although the modular group did marginally better on their undergraduate course. On the other hand, Taverner and Wright (1997) found that while students on modular courses were possibly more successful than those on linear ones (up to half a grade difference, probably because students were re-sitting modules on which they received an N or a U) the teaching they experienced was more didactic. McClune (2001) investigated the effect of maturity upon students' performances. A level physics students performed appreciably better on the same assessments in the second year of sixth form compared with the first. He explained these findings by reference to a maturity factor – students were able to consolidate their knowledge over time. Hayward and McNicholl (2007), in an analysis of modularisation of science A levels, believed that the benefits of modularisation – short term goals, regular feedback, flexibility, improved progression – were outweighed by the drawbacks – fragmentation of knowledge and an increasingly mechanical and instrumental view of teaching and learning, exacerbated by national accountability measures. They argued that students understand science on a more shallow and superficial level in a modular system. Since no randomised control trial has been conducted comparing the outcomes of students sitting modular and linear examinations, all of the studies that have sought to make this comparison suffer from the confounding factor of teacher and student choice. Those who opt to take a modular examination may be better suited to that mode of study.

The Smith Report into mathematics of 2004 claimed that modularisation was felt by many respondents to hinder the development of the learning and understanding of mathematics at this level. A concern was that too much time was devoted to examinations and preparing for examinations – 'teaching to the test'– and that this was at the expense of understanding of the subject itself. Many identified the problem as splitting of the subject matter of A level mathematics into six separately examined modules. This was seen as having the effect of splintering the unity and connectedness of the mathematics to be learned at this level (Smith Report, 2004, p.93-94).

Students' opinions varied, according to Hodgson and Spours (2005), with some believing that taking modules in their first year gave them greater control over their learning and others simply not feeling ready for early assessment. However, they found students broadly positive, appreciating that they could revise and take examinations one at a time and re-sit if things went badly.

Another review found that students largely favoured modularisation, believing that it made learning easier because of the small units, the clearer content and the immediacy of testing after learning the material (Alpha Plus, 2012). They also liked being able to take examinations throughout the course of study, finding that approach less stressful than taking all of the assessments at the end.

Chapter 4 Examination structure research

Claims regarding the advantages and disadvantages of linear assessment

The literature search (Appendix D) indicated that a range of claims about the relative benefits of linear and modular examination structures have been made by researchers (Appendix E). These relate to effects of examination structure upon learning, outcomes and progression, teaching, the student experience, student attitudes and system-level issues (Table 4.1). Below we turn to empirical findings on examination structure, but for now we outline the claims that are made for them. Note that there is rarely empirical evidence for these claims and where empirical evidence exists, it is typically scant.

Effects upon learning

A perennial concern is that examinations *in general* produce short-term learning goals, which induces instrumental motivation in students and thwarts deep learning and long-term retention. These effects are considered to be more severe with modular examinations. A coherent learning experience in which connections are made across a subject area is also said to be undermined by modular examinations, which could disrupt understanding. After all, to understand a subject in depth is to comprehend the underlying structure of the discipline. Subjects which require extended writing for learning purposes are also said to benefit from linear assessment, largely because there is greater time to develop those skills. Modular examinations, though, can allow students to master a topic before moving on to the next topic and are better aligned with 'testing when ready' philosophy. Students need not necessarily sit a modular assessment until they are ready to achieve on that test.

Outcomes and progression

Although the claims relating to learning tend to favour linear examinations, it has also been noted that modular examinations can improve achievement and result in higher grades. In part, higher outcomes arise because re-sits permit students the opportunity to work harder, try again and improve their grades. The counter argument is that modular examinations are 'dumbing down' examinations, are not as challenging, and that the rises in outcomes associated with them are 'grade inflation.' Linear exams (at A level) supposedly prepare students better for university and are more trusted by universities and employers.

Teaching

Modular exams at GCSE occur so frequently that they are disruptive to teaching. Not only are students out of the classroom for examinations, exam preparation consumes more teaching time and there is therefore less focus upon the subject content in lessons. On the positive side, it is sometimes claimed that modular examinations allow more flexibility. Additionally, because teaching can be presented in discrete chunks, teacher expertise can be organised and scheduled more effectively, allowing teachers to specialise in topics.

Student experience

High-stakes examinations are stressful for students, while modular examinations reduce the stakes and therefore can reduce stress. For some students though, the constant pressure to perform and the workload associated with frequent examinations can also prove stressful. Setting targets and getting feedback on those in the short-term could be motivational. Equally, students know how much they need to do in the remaining modules and re-sits to attain particular grades, which is important for tracking their progress. Occasionally, this can backfire, with students realising that they do not have to achieve much at all in the final examination to reach their target grade. Nonetheless, uncertainty is reduced with modular assessment and the shorter examinations with less subject content can be easier to prepare for. Despite this, some students experience an education that is dominated by examinations in modular schemes, but that need not necessarily be the case because they do not have to sit every examination series available; modular schemes allow flexibility. Staged assessment in modular schemes might disadvantage students because students are less mature when they sit early modules. This could be more important for some subjects than others, such as English.

Student attitudes

Good modular results can ease the pressure on students, but weak results could be demotivating. Effort might be limited on initial modules because students know that they can re-sit. The provision of clearly defined curricula and targets associated with modular schemes can be helpful for students; some argue that this is particularly good for less independent learners. A big potential problem with this approach, though, is that it can increase instrumental attitudes to learning. Again, it is not that instrumental approaches to learning are *only* associated with modular schemes, but that this is claimed to be more so with modular exams than with linear.

System challenges

Modular examination systems are more expensive because there are more examinations, with all of the associated administrative costs. They are complex schemes too because there can be many routes through a qualification. Such complexity has resulted in significant problems in the theory and practice of setting examination standards, as discussed previously.

Table 4.1 Relative benefits of linear and modular exams as claimed in the research literature

	Linear exams		Modular exams	
	Benefits	*Disadvantages*	*Benefits*	*Disadvantages*
Learning	long-term retention of information		testing when ready	less subject coherence
	foster depth of learning		master a topic before moving on	reduced time for extended writing
	better development of subject-specific skills (due to additional time)			
	better understanding of subjects			
Outcomes & progression	aggregation effects more advantageous to top students		higher grades	grade inflation
	prepare students for A level study		re-sits allow improvements	
	trusted by universities and employers			
Teaching	more time for teaching due to less frequent exams		allow more flexibility	
	more exam preparation		better allocation of staffing and resources	
			present content in manageable chunks	exams disruptive
	time for focus upon subject content			

	Linear exams		Modular exams	
	Benefits	***Disadvantages***	***Benefits***	***Disadvantages***
Student experience	stress lower – exam load	stress – high stakes	allow students to track their progress	constant workload
	maturity when taking exam		frequent feedback helps students improve	
	course not dominated by assessment		short-term targets beneficial	
			easier to revise	
			flexibility to take either examination structure	
			reduced uncertainty	
Student attitudes			good for less independent learners	poor modular results demotivate students
				increase instrumental attitudes
				re-sits mean students do not have to try so hard the first time
System challenges				setting standards (early modules, between series, etc.)
				expense
				complexity of administration

Evidence on the effects of modularity

Re-sitting, with special attention to mathematics and English

Simply re-sitting an examination might be thought to have no benefits to the candidate unless they have revised harder. However, all measurements, not just examinations, have some degree of error associated with them. Since students keep their best grade to date for a module when they re-sit, there is opportunity cost (in time, money, etc.) but not a direct risk to their examination results. It follows logically too that students do not have the penalty of error on a re-sit when it is not in their favour but they benefit from error when it increases their grade. In other words, the application of the re-sit rules means that the treatment of error is biased in students' favour. The higher the number of re-sits students take, the higher the likelihood of a 'false positive'; that is, the student being given a grade that they do not deserve. Using simulation methodology, Wheadon (2010) showed that after four re-sits, for example, only 44 per cent of students would receive the grade that they deserved. However, a low proportion of students actually sat the same examination four times, or re-sat modules for the same qualification in A levels or GCSEs, as we see below.

Re-testing on cognitive measures

There is a wider literature on the effects of re-testing on psychological, cognitive ability measures, which looks at practice effects. These do not involve test preparation as we might expect for GCSEs; only multiple exposure to the test. A meta-analysis of re-testing effects showed a small, positive effect (Cohen's d = .26) of test practice (Hausknecht et al., 2007). Re-testing gains reduced with each subsequent opportunity, but were in evidence for up to six re-sits, and people with higher cognitive ability benefitted more from re-testing (Kulik et al., 1984; Arendasy & Somer, 2013). Analytical and quantitative tasks show higher re-testing effects than verbal tasks (Hausknecht et al., 2007). Recent research indicates that re-testing effects are domain-specific, as opposed to showing an increase in general intelligence, or g (Arendasy & Somer, 2013). As students are prepared for GCSEs through teaching on the curriculum and exposure to prior examples of the tests, the findings relating to cognitive ability tests might not generalise to curriculum-related educational assessments. However, it is worth knowing that in the absence of a curriculum and preparation, re-testing benefits the high ability students and has most impact upon quantitative tasks and domain-specific content rather than more generic thinking skills.

Re-sitting at A level

Much of the evidence relating to re-sits concerns A levels. Re-sitting or re-taking elements of a course – module examinations and/or coursework – became a serious issue after the introduction of fully unitised A levels in 2000. If qualifications are primarily assessed in a linear fashion at the end of a course then candidates largely need to re-take assessments for entire qualifications (the exception has been coursework, the grades of which could generally be brought forward for the lifetime of the syllabus), which naturally inhibits the number of re-sits. Curriculum 2000 changed that because all A levels became modularised with three units of AS qualifications complemented by three A2 units. While the

rationale for allowing re-sits at A level was not publicised by either the government or the regulator – at that time a division of QCA – Sutch and Wilson (2013) posited that because A levels are such high-stakes qualifications, both for institutions and for candidates themselves, re-sitting provides opportunities for candidates who believe that their unit results were lower than they would have expected or believe themselves capable of rectifying the situation. This notion of getting the grade one deserves through further work and re-sits was reinforced by Ron Dearing in 2006 and Ken Boston, then CEO of QCA, in 2007 (MacLeod, 2007 in Sutch & Wilson, 2013) during the initial A level development. It was argued that re-sitting and modular assessment were essentially complementary functions, in which results from an early sitting could act as a formative assessment, providing candidates with feedback as well as the motivation to do better next time (Gray, 2002 in Sutch & Wilson, 2013). Originally, re-sits were limited to one per unit, but after the first years of Curriculum 2000, when both re-sit and 'cashing in' – certification – rules were deemed to be overly complicated and unmanageable, QCA allowed unlimited re-sitting of A level modules, under the assumption that very few candidates would re-sit units more than once.

While a QCA study (2007c) provided evidence that the majority of re-sitting candidates did so only once per unit, the notion of a 're-sit culture' quickly took on a life of its own when it was believed that many candidates were re-taking units in which they had done perfectly well, simply to be able to bank more UMS (Uniform Mark Scale) marks in earlier units to improve overall A level grades because the best of the outcomes was counted (de Waal, 2009; Higton et al., 2012). QCA found in 2008 that between two thirds and three quarters of candidates included in the sample they investigated had re-sat at least one unit (DfE, 2010). AQA reported that more than half of A level candidates who took their examinations had re-sat at least one unit and a quarter re-sat two or more units (AQA, 2012b). If re-sits were disregarded from examination grading, almost five per cent fewer candidates would have achieved a grade A.

According to Scott (2012), because A levels are such high stakes both for schools and colleges and for candidates, students were tempted to engage in questionable re-sitting practices once unlimited re-sits were allowed. She feared that some re-sitting students got more help and support than others, particularly those from independent schools and colleges. Over-reliance on second chances; increasingly narrowed focus on unit grades rather than on learning; and lessening of learner autonomy due to teaching to the test (common, of course, to all high-stakes assessments, not just to re-sits) were also issues.

Gill and Suto (2012) argued that the situation was more nuanced than candidates simply sitting the easier and earlier AS modules for practice and mark banking, arguing that re-sits were more strategic and very few candidates first took units for practice. Once unit numbers were decreased from six to four for most A levels in 2008 and the A* grade introduced, the gaining of which depended in large part on achievement in the less often re-sat A2 units, re-sits decreased, reinforced by the abandonment of January sittings in 2013 (no January examination was available in 2014). New A levels to be first taught in

2015 were linear in fashion, so to improve results candidates had to retake entire qualifications, with the exception of coursework for which results could be carried forward.

Re-sitting at GCSE

For GCSEs, the re-sitting situation is more subtle. Rather than allowing unlimited re-sits as for the A level units, the GCSE qualifications criteria (QCA, 2007a; QCA, 2008) that regulated assessment when the examination boards began to offer modularised GCSEs in earnest only allowed each unit to be re-sat once, with the better result counting, and required that candidates take at least 40 per cent of assessments at the end of the course. This may have inhibited re-sits for practice. Vidal Rodeiro and Nádas (2012) conjectured, however, that it is possible that the existence of re-sits might mean students take their first attempt less seriously than they might otherwise.

In the two most high-stakes subjects, mathematics and English, Vidal Rodeiro and Nádas (2012) found a mixed picture of achievement for modular route re-sitting candidates. The candidates they studied certificated their qualifications in 2009, before the wholesale introduction of modular GCSEs. In mathematics most students on the modular route who re-sat units did better on the re-sit than the first sitting. A relatively high percentage of candidates re-sat units in mathematics; one third of candidates re-sat one unit and under two per cent re-sat more than three units. For all 10 units that they analysed, the average marks obtained in the second attempt were significantly higher than in the first one. Between 30 per cent and 50 per cent of the re-sits led to an improvement in the unit's grade and in less than 10 per cent of the cases the grade was worse. Better grades on individual units, of course, do not necessarily mean better overall qualification grades.

The findings were similar for English. Here, few students re-sat units,[20] but for those who did, in more than 40 per cent of the cases the grade was better in the second attempt and led to an improvement in the unit grade. However, a sizeable minority of re-sits for the externally marked units (between 15 per cent and 22 per cent) resulted in lower grades. And, as a harbinger of the summer of 2012 (see Appendix B), for the coursework units in 98 per cent of the cases a re-sit resulted in the same or better grade. According to Vidal Rodeiro and Nádas (2012) the benefits of re-sitting were clear, although they included the caveat that some of the candidates might originally have been entered for a unit too early or for motivational purposes. They argued that allowing re-sits could mean that students were learning more and could be of benefit to weaker candidates.

Early and multiple entries at GCSE

Another issue associated with the modular GCSEs is increased early and multiple entry and certification, especially in English and mathematics, although there is evidence that this phenomenon began before 2009. Since units could only be re-sat once and (since 2007 the 40 per cent terminal assessment rule has been in place) having assessment opportunities at least four times over Key Stage 4 (and,

[20] The article was written before English became fully modularised in 2010-2012.

for schools that start students on GCSEs in Year 9 six times) can allow students who sit units early to certificate their achievement and potentially sit the entire qualification again before year 12. An Ofqual (2013a) study found that early entry for English GCSEs started rapidly increasing in 2008 and for mathematics in 2009. By 2011 over 200,000 students took their GCSEs early in each subject, in contrast to 9,000 (English) and 24,000 (mathematics) in 2005. Some schools entered all of their students early.

Schools were judged by how many of their students reach a grade C or above in English and mathematics,[21] although this changed to some extent with the introduction in 2016 of judgments based on candidates' achievements across their 'best eight' GCSEs (and equivalent qualifications). Schools can hedge their bets by entering some or all of their students for certification in Year 10, hoping they will achieve a grade C or better and re-entering them if they do not. Some schools, especially in mathematics where the subject content rules mean that examination board specifications are not dissimilar (unlike English where set texts can differ dramatically), entered students for GCSEs from more than one examination board (ACME, 2011; HoC, 2013a), again to increase the chances that students will get lucky and achieve a grade C on at least one of them.

Ofsted (2013) investigated the impact of early entry and their main findings were:

- That while early entry for English and mathematics had increased dramatically, its use in other subjects was limited, doubtless due to accountability pressures.
- While in the past early entry was more likely used for the most able students, in some cases allowing them to begin their A levels early, schools in 2013 were entering students of all abilities.
- Candidates who took English and mathematics early were less likely to achieve A or A* than those who did not, possibly as a by-product of the finding above.
- Students in some schools no longer had English and mathematics classes once they had achieved their target grades, even if they had achieved below their potential.
- Schools with a higher achieving student body were less likely to enter candidates early and schools that had been rated satisfactory or lower by Ofsted appeared to using early entry more extensively.
- While some schools used early entry highly effectively, others did not, with little positive impact on standards.
- There was worry on the part of subject specialists and advisers that early entry weakened the foundation for post-16 study because in some schools it limited higher grade achievement, constrained the curriculum and restricted post-Key Stage 4 choices.

[21] Now grade 5 in GCSE English and Mathematics

As discussed earlier, the Coalition Government made clear from the outset their view that the modularisation of GCSEs had been a mistake (DfE, 2010). The government stated that GCSEs were too small to be taken 'sensibly' in units across two years and also that the assessment burden was too great, with schools entering candidates in Years 9 and 10 as well as 11. Echoing the Ofsted report, the then secretary of state, Michael Gove, described early entry as harmful to the interests of students (DfE press release, 2013).[22] Concerned that in the summer of 2013, twenty-three per cent of mathematics entries and 10 per cent of English entries had been early (an increase of 39 per cent from 2012 to 2013), he stated that early entrants performed worse overall than those who took their exams later, even taking re-sits into account. Concerned that schools were entering candidates before they were ready in order to bank C grades, he pointed out that for some of those candidates Key Stage 2 performance would suggest that had they carried on until the end of Year 11 they might have achieved a grade B or better.

To inhibit early entry for qualifications, since the autumn of 2013 results from a student's first entry in English, mathematics, science, history, geography and modern foreign languages count toward their school's performance tables (other subjects followed in 2016). Individual students might benefit from re-taking a GCSE, of course, but some schools might not be willing to engage in this risky (for the school) strategy.

Gill (2013) did not believe that the changes to GCSEs that began in 2012 would have as great an effect on early entry as the government hoped. Although candidates have to take all assessments at once, they can still do so in the summer of Year 10. He thought it was possible that more candidates would certificate early because they cannot take individual units early and re-sit if necessary; instead they may sit all their examinations early to be able to re-take the entire qualification before the end of Key Stage 4. However, the changes to performance table rules could inhibit schools from doing this and in the summer of 2014 there was a 40 per cent drop in the number of entries by pupils aged 15 and under (JCQ, 2014). Those 15-year-olds who entered early were more likely to get higher grades than in the immediately preceding years, perhaps because schools had reverted to their early pattern of only entering highly achieving students early. Early entry for mathematics dropped from 170,000 entries to 39,000.

Wales also changed their performance tables so that from 2019 only the first GCSE result would count.[23] This followed a Qualifications Wales report on the effects of early entry on teacher's decisions with regard to examination entry (Sperring, Davey, Jones and Anderson, 2017). The prime rationale for the change was the limiting effect upon young people's achievements, as teachers were focused upon gaining as many grade C results as possible for the performance tables used to evaluate school effectiveness.

[22] https://www.gov.uk/government/news/changes-to-early-entry-at-gcse
[23] https://gov.wales/changes-early-entry-exams-announced-kirsty-williams

Chapter 5 Economic impact of the structure change

One of the ways that the change in qualification structure could affect the education system is in economic terms: for schools, examination boards, and the system more broadly. Further, as GCSEs are offered in a competitive market in England, changes in their structure affect the properties of the market itself as well as behaviour within it.

The market

England's examinations market is a regulated oligopoly (Jones, 2011) with a few organisations dominating the market. Apart from in mathematics, no examination board chose to offer a linear version of the GCSE when the 2007 reforms were carried out. This represented a change in the marketplace, as modular GCSEs had previously been offered as alternatives to linear GCSEs, unless prohibited by the regulator. Clearly modular GCSEs were attractive to the market and/or the examination boards. Of course, a modular qualification can be taken in a linear manner and examination boards may have judged that a modular product would be a less risky, more flexible offer. Indeed, a sizeable proportion of schools took this pseudo-linear option.

It is also worth noting that the context of the 2007 reforms was such that unitisation fitted with wider policy directives: the qualifications frameworks of the time, the Diploma and the modular A levels. The more recent 2015 move to entirely linear GCSEs, however, was prescribed by the government in response to concerns that modularisation had a negative impact on teaching and learning and undermined standards. The new GCSEs were reformed to tight criteria which were designed to ensure comparability between GCSEs offered by different examination boards. The intention was that examination boards would be more likely to compete on the services offered to schools rather than on the standard of the qualification.

Qualification fees

Has the recent move to linear GCSEs impacted the fees charged by examination boards? This question is difficult to answer. The change from modular to linear structure did not occur in isolation. There were a number of other changes made at the same time, including a move to more examinations rather than teacher assessment which increased the costs to the examination board of delivering of the qualification. For example, in GCSE English the reformed GCSE comprises two examination assessments along with a separate teacher-assessed grade for spoken language. The legacy qualifications, on the other hand, composed 40 per cent teacher assessment. There was also an increase in the volume of subject content in GCSE mathematics. This is reflected in the four-and-a- half- hour examined assessment compared to the legacy qualifications which had between three quarters of an hour to an hour less of examination time. Examination marking is a key cost driver in qualification design.

Further, examination boards cross-subsidise their qualification offer. The fees rarely reflect the true cost of running any individual qualification. Indeed, some boards charge a flat rate across their GCSE suite, probably on the basis that the cost of a qualification ought not to influence school and college choices of curriculum. Therefore, changes elsewhere in the qualification landscape, for example the coincident introduction of examinations into qualifications such as BTEC First Award or BTEC Tech Award into performance table qualifications, may indirectly influence the pricing of GCSEs.[24] Moreover, some examination boards have significant other commercial interests such as in publishing. The success of these interests may also influence qualification pricing.

A final complicating factor is that examination boards need to recoup the costs of qualification reform. They engage in what is known as an 'invest and then harvest' strategy (Frontier Economics, 2015). Most schools and colleges tend to stick with a provider of qualifications during non-reform years, only considering a change at the time of reform (Ofqual, 2015a). Hence, examination boards invest and compete strongly to attract schools at times of reform. They then relax their efforts once that has been achieved. Interestingly, Ofqual (2015a) reported that in choosing provider, the vast majority of schools and colleges did not consider fees, rather they focused on the match with their students' needs, the content and the style of the assessment.

The 'invest and harvest' phrase can also be applied to examination board pricing strategy across the qualification life cycle. The significant costs incurred during qualification reform are borne by the examination boards in the short term, to be recovered over the anticipated life of the reformed qualification. Without large surpluses in the non-reform years, examination boards would struggle due to the costs of reform and marketing and would have to cross-fund their activities from other sources.

Table 5.1 provides the average prices between 2015 and 2018 for some key GCSEs. The GCSE reforms are being introduced in waves. The shaded cells show the prices for the legacy qualifications and the non-shaded cells the prices for the reformed linear qualifications. The changes in average prices from legacy to reformed GCSEs vary from £3.81 for GCSE English to £1.48 for GCSE French. In general, but not exclusively, the annual increase in average price during the change from legacy to reformed qualifications is greater than the annual increase in price during stable periods. The greatest annual average price increase during this time was £1.93 for GCSE history between 2015 and 2016. However, it would be unwise to conclude that linear GCSEs are inherently more costly than modular GCSEs for the reasons set out above. In general, price increases were in excess of changes in the retail price index during this period which might be expected given the costs of reform.

[24] https://www.gov.uk/government/news/performance-table-reform-and-transparency-will-raise-standards-and-end-perverse-incentives

Table 5.1 Average GCSE fees for AQA, Edexcel, OCR and WJEC weighted by entries

Average price (weighted by volume)	2015	2016	2017	2018
GCSE English	30.44	31.45	35.26	36.61
GCSE English literature	29.71	31.48	33.90	35.28
GCSE maths	31.54	31.47	33.81	36.24
GCSE science	31.02	32.39	33.47	35.17
GCSE French	31.19	31.79	32.72	34.20
GCSE history	30.55	32.48	33.34	36.42
GCSE art and design	30.44	31.37	32.88	35.23
Annual RPI	2.4%	1.0%	1.8%	3.6%

Cost to schools and colleges

Rising examination expenditure by schools and colleges has been a source of concern that has ebbed and flowed but was prominent during the period of entirely modular GCSEs (see for example, the Association of Colleges, 2014; Ofqual, 2015a). However, the cost of GCSEs to schools and colleges is not only determined by the fees for the qualifications. Changes in the number of GCSEs sat by pupils, re-sitting of either units or whole qualifications, double entry and late entry, and reviews of marking (appeals) are major factors influencing expenditure (Ofqual, 2015a). These all peaked during this period and are likely to explain a great deal of the concern regarding rising examination costs.

To give a sense of the potential impact of multiple entry on cost, it is helpful to consider GCSE mathematics during the academic year 2012-13. In the Joint Council for Qualification's (JCQ) presentation to the press of the 2013 provisional results, it was explained that whilst the 16-year-old cohort numbered 678,000, GCSE mathematics entries were nearly double that figure, at 1,326,003 across the four available examination series – November, January, March, June. This multiple entry will have had significant cost implications for schools and colleges (DfE, 2013; Taylor, 2018).

We have calculated the number of students who certificated in two or more GCSEs or International GCSEs in mathematics and English between 2013 and 2015 (Table 5.2). As can be seen, a substantial amount of double entry occurred in this period. This practice led to fears that pupils were entered before they were ready and 'banking' a grade C where their performance at Key Stage 2 would suggest that if they had continued to study the subject and taken the GCSE at the end of year 11 they could have achieved a higher grade. As discussed previously, in response, in 2013 the government announced changes to accountability measures such that a pupil's first entry into a GCSE was that which counts toward a school's performance.[25] Early and multiple entry into GCSEs immediately decreased.

[25] https://www.gov.uk/government/news/changes-to-early-entry-at-gcse

Table 5.2 No. of students certificating in two or more GCSEs/International GCSEs in maths and English

Students certificating in 2+ GCSEs or International GCSEs	2013	2014	2015
English	86,129	17,762	10,319
Maths	292,118	101,667	24,176

It is clear that wider policy can dramatically impact on cost but in principle, the move from modular to linear qualification structure should be beneficial to schools' costs. In particular because of the associated reduction in costs of re-sits of individual modules. To give a sense of the cost of re-sitting modules, consider the work of Vidal Rodeiro and Nádas (2012) who analysed GCSE English and mathematics data from one of the examination boards, OCR. They reported that among all the students who certificated in 2009, there were 1,508 (3.58%) students who re-sat one of the five GCSE English units, 1,095 (2.60%) students who re-sat two units, and 1,630 (3.87%) students who re-sat three or more units. In English, just below 90% of the students did not re-sit any unit. This would represent a cost of £85,880 to the system if a price of £10 per GCSE module re-sit is assumed. Whilst this is a small figure compared with the overall cost to the country of examinations, only one syllabus is considered here.

In mathematics the picture was more extreme. There were 17,089 (31.73%) students who re-sat one of the 10 GCSE mathematics units, 9,752 (18.11%) students who re-sat two units and 1,178 (2.19%) students who re-sat three or more units. In mathematics only 47.97 per cent of students did not re-sit any unit. For this syllabus, the costs would be approximately £400,000, again assuming a fee of £10 per GCSE. Therefore, it is safe to assume that the costs of re-sitting were non-negligible.

Another cost reduction is that associated with late entries which are less likely under the simpler linear structure. Indeed, Ofqual (2017a) reported that in 2017 the number of late GCSE entries dropped by just under 10 per cent from 485,785 to 437,945 but continued to represent approximately three per cent of all entries. As the reformed GCSEs are rolled out the number of late entries is expected to drop further. Late entries are discouraged by the examination boards and can cost double the price of a standard entry so this will represent a considerable saving to schools and colleges.

Wider costs to schools and colleges

In 2013 Ofqual published their regulatory impact assessment of the 2015 reforms to GCSEs, including the move to linearity. Interviewed school representatives believed that for them the impact of the proposed reforms would be cost neutral. While they expected to incur increased costs for new resources, materials and teacher training, they thought that this would be offset by lower expenditure on, for example, examination fees as well as savings in teacher time by the removal of controlled assessment.

The reduction in examination fees was expected to come from fewer entries reflecting reductions in the number of modular re-sits and fewer late entries

stemming from more straightforward administration and the reduction in tiered subjects. Some schools suggested that students would also take fewer GCSEs because they would be more challenging and some would require longer teaching time.

Schools expected that the reforms would make examinations administration simpler. However, some schools and examination officers were concerned that the concentrated summer examination series would affect the availability of examination rooms and would create pressure on the examination office during this period. Potential implications for school management information systems were identified as another cost.

From the evidence discussed above, there is no reason to believe that linear GCSEs are more expensive than modular GCSEs, quite the opposite. However, wider policies than the qualification structure have more impact upon examination costs.

Chapter 6 Pupil level quantitative data analyses

Data selection and the National Pupil Database

Background

The National Pupil Database (NPD) holds a wide range of information about students who attend schools and colleges in England. It combines the student-level examination results with information on student and school characteristics. Over and above the information published by the JCQ,[26] it details prior educational attainment, gender and proxy measures of socio-economic status for each student. The data available for evaluation of the effect of modular and linear assessment are taken primarily from the results section of the NPD. This section is used to create the school performance tables. The data analysed cover the academic years 2006/2007 to 2013/2014 for GCSE students. Data from GCSE English, mathematics and science form the basis of the analysis. These GCSEs are all EBacc subjects[27] and therefore the entry is large – close to that of the full age cohort.

A separate NPD dataset contains A level results for all students entered between 2006/2007 and 2014/2015. These data are combined with the GCSE data to consider the longitudinal effect of modular and linear assessment at GCSE.

Data preparation

For some students, the NPD contains multiple records. Students often re-sit their GCSEs in key subjects, entering either the same specification or choosing an alternative in subsequent attempts. Multiple entries in the same subject area in the same series are also evident. Sometimes this is because a student has moved school midway through the course but sometimes it is more related to how the data are recorded for use in the performance tables.

Multiple records are particularly apparent in the winter series between November 2010 and March 2013. In these years, there were three separate winter series which were all coded on the database as winter (November, January & March). The majority of winter entries for English were made in November and, for science, in January. However, for mathematics, the winter entries were spread across November and March. Some students were entered in both series. In summer 2012, the JCQ press release stated:

> The results show a year-on-year decrease of 12.6 per cent in the number of entries for Mathematics in summer 2012. When the whole academic year is considered, the number of entries increases by 17.6 per cent. The main factor in this variation is an additional examination opportunity in March 2012 (JCQ, 2012)

[26] https://www.jcq.org.uk/examination-results/gcses

[27] The EBacc (English Baccalaureate) is a school performance measure. The EBacc entry and attainment measures allow people to see how many pupils are entered for and get a grade C or above in the core academic subjects at Key Stage 4 in any government-funded school. The EBacc is made up of English, mathematics, the sciences, a language and history or geography. The EBacc is made up of English, mathematics, the sciences, a language and history or geography.

The decision on how to deal with multiple records influences the interpretation of any subsequent analysis. The NPD requires some manipulation to be suitable for use in the evaluation of the effects of modular and linear assessment. Steps were taken to format the data (Appendix J) and are illustrated in the flow diagram included in Appendix F.

Subsequently, two datasets were created. The first was designed to investigate the effect of teaching a modular or linear course up until the point of first certification (D1). The second was designed to allow exploration of the effect of assessment route combined with recertification opportunities. Recertification is when a new grade is awarded. It is distinct from re-sitting as, depending on the policy in a given year, students could re-sit units within a qualification a variable number of times prior to certificating.

The first certificated attempt (D1)

The first certificated attempt at a GCSE in the subject of concern was retained for each student. For modular specifications, this first attempt may have included multiple sittings of the constituent units. For linear specifications, it represented the first formal examination. Whatever the specification, the GCSE may have been taken at the traditional age of 16 or as an early entry. The distribution of entries by subject, academic year and assessment route is given in Figure 6.1.

Results from the first certificated attempt are influenced by a number of significant factors. For instance, the approach to teaching may differ dependent upon whether the specification is linear or modular. Entry patterns might also vary. Furthermore, the hidden effect of re-sitting units in a modular specification before first certification is liable to influence the results. Pre-certification module re-sitting is a factor not recorded on the NPD. Thus, any analysis should be interpreted with these limitations in mind.

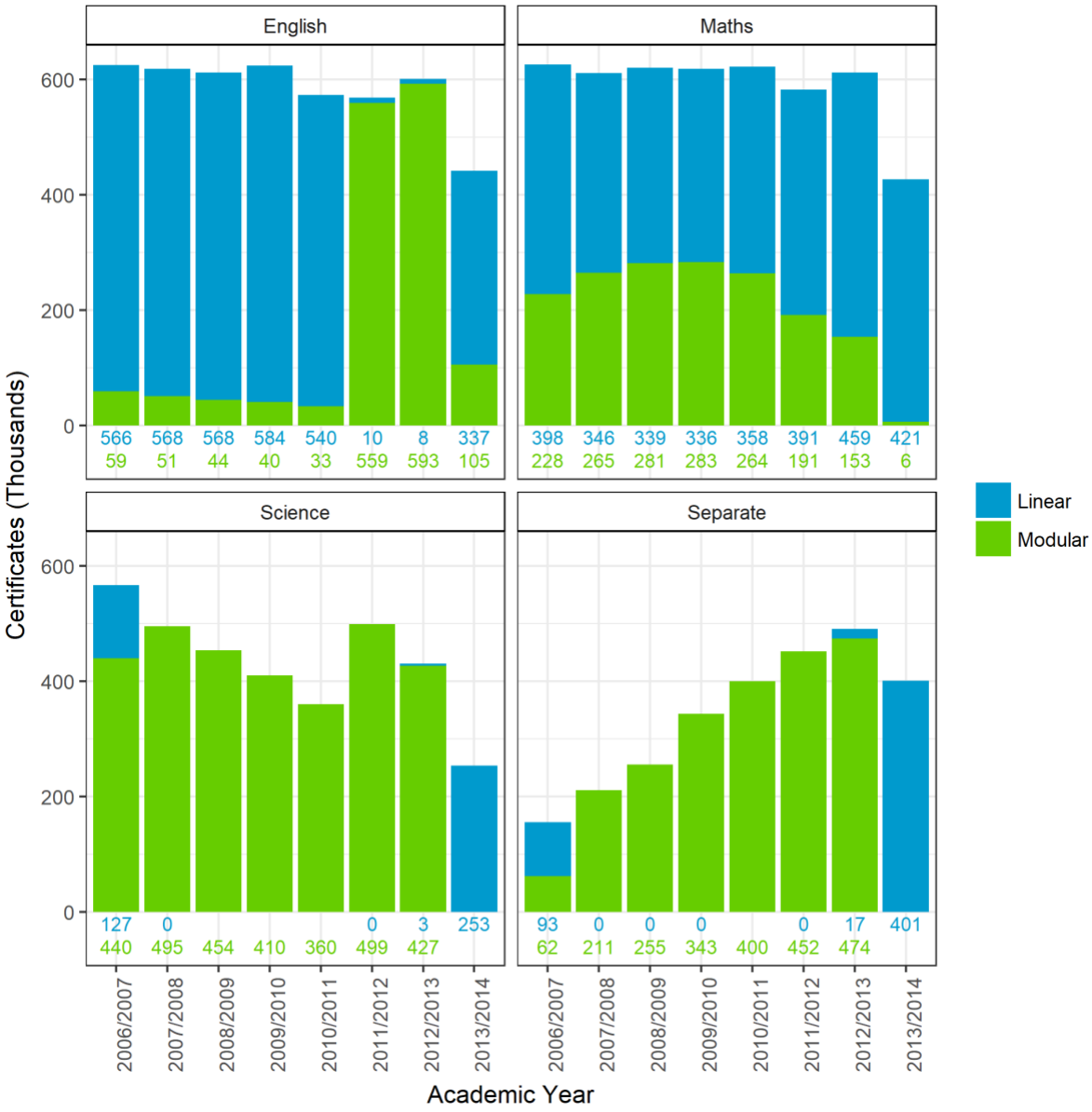


Figure 6.1 The distribution of modular and linear GCSE results at students' first certification in a subject

The best certificated attempt in each academic year (D2)

For each subject, the best certificated attempt in any given academic year was retained for each student. For both modular and linear specifications, the dataset included multiple entries for given students reflecting year-on-year recertification opportunities. The distribution of entries by subject, academic year and assessment route is given in Figure 6.2. Once again, the best attempt in any given academic year is influenced by a number of significant factors. While results are undoubtedly affected by teaching and assessment route, they are also influenced by re-sitting and re-certification opportunities.

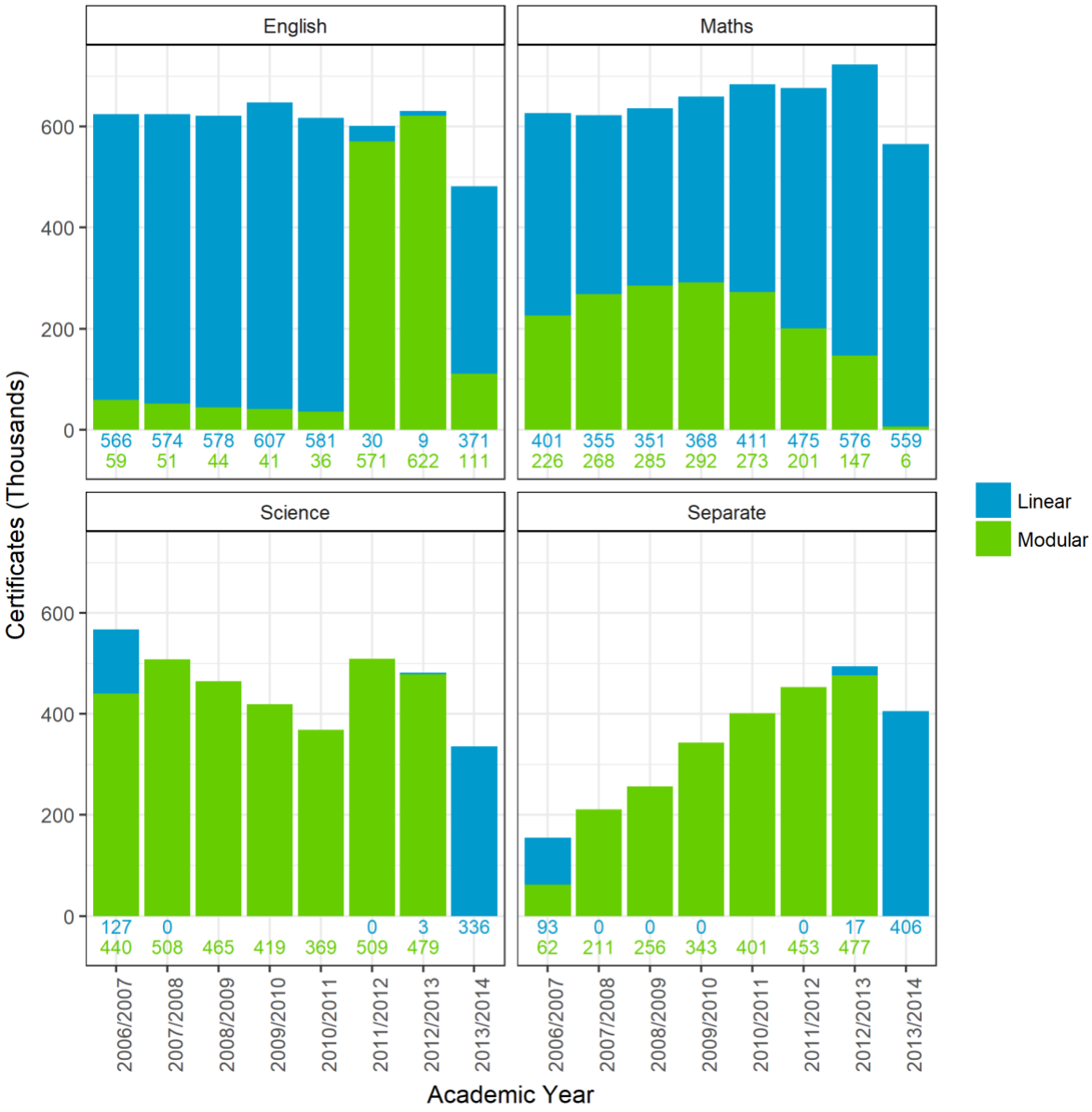


Figure 6.2 The distribution of students' best modular and linear GCSE results a given academic year

Comparisons with the JCQ data

To provide some context and validation of the data from the NPD, Table 6.1 allows comparison with the JCQ results statistics. For the summer series, the provisional GCSE results for England are reported on the JCQ website.[28] These statistics are compiled at the time of issue of results so do not include grade changes due to post-result enquiries. The winter statistics are not readily available. Therefore, comparisons are for summer only and are made against D2 described above.

[28] https://www.jcq.org.uk/examination-results/gcses

The patterns of entry in the JCQ data are broadly the same as that seen in the NPD data illustrated in Figure 6.2. On the whole the JCQ data include a greater number of certificates. This is liable to have occurred for one of three reasons. First, the NPD data (D2) do not include every result in a series, merely the best. Second, the NPD dataset records student results against their school year and therefore, particularly for the later years, some early entry students would be omitted. Third, any NPD record for which there was no modular or linear flag was omitted.

The two occasions on which the NPD data are significantly higher than the JCQ data coincide with years when there were considerably greater numbers of winter entries for the subject. Change in specifications, movement to International GCSEs and change in recording of the statistics are all liable to have contributed to this difference. Thus, the comparisons with the JCQ data are reassuring with respect to the NPD analyses.

Table 6.1 Comparison of D2 with the summer data available from JCQ (entries in thousands)

	English			Maths			Science			Separate		
	JCQ	NPD	% Diff	JCQ	NPD	% Diff	JCQ	NPD	% Diff	JCQ	NPD	% Diff
2007	667	627	-6.0	693	628	-9.4	579	569	-1.7	163	156	-4.3
2008	658	626	-4.9	672	624	-7.1	501	509	1.6	217	211	-2.8
2009	639	624	-2.3	690	637	-7.7	459	466	1.5	262	256	-2.3
2010	643	649	0.9	698	660	-5.4	417	420	0.7	349	345	-1.1
2011	590	617	4.6	709	685	-3.4	372	369	-0.8	405	401	-1.0
2012	612	601	-1.8	613	677	10.4	506	510	0.8	456	454	-0.4
2013	666	633	-5.0	695	724	4.2	421	483	14.7	474	495	4.4
2014	454	483	6.4	666	565	-15.2	347	336	-3.2	389	406	4.4

Covariates and missing data

Missing data – the first certificated attempt (D1)

The formatted dataset includes 4.67 million GCSE English results, 4.73 million GCSE mathematics results, 3.47 million science results and 2.72 million separate science results (see Appendix F, D1.7). It represents a total of 5.14 million students taking examinations in any or all of the subjects of concern. However, in the NPD, not all fields are completely populated and therefore the effective size of the dataset is somewhat smaller. For the purposes of modelling the effect of assessment route, complete data are needed for all variables in the model.

Table 6.2 lists potential covariates which might influence GCSE performance and it details the extent to which the data are complete for these covariates. The linear/modular indicator, academic year, examination series, year-end age, gender and school type are complete for all records in the dataset. All the other variables include missing values. The missing values affect the number of

records that can be used to build a model. With the exception of the prior attainment measure for separate science students, all the variables have less than 11% missing values.

Any model controlling for the linear/modular indicator, examination series, year-end age, gender, school type, IDACI (Income Deprivation Affecting Children Index), the free school meals indicator, SEN (special educational needs) indicators and prior attainment will be based on 89.2% of the available GCSE English results, 90.3% of the GCSE mathematics results, 90.9% of the GCSE science results and 83.0% of the separate science results (see Appendix F, D1.8).

Table 6.2 The percentage of missing values for each of the potential covariates in the model

Covariate	% Missing			
	English	Maths	Science	Separate
Linear/modular indicator	0.0	0.0	0.0	0.0
Academic year	0.0	0.0	0.0	0.0
Examination Series	0.0	0.0	0.0	0.0
Year-end age	0.0	0.0	0.0	0.0
Gender	0.0	0.0	0.0	0.0
School type	0.0	0.0	0.0	0.0
IDACI	0.1	0.1	0.1	0.0
Free school meals	2.4	1.8	1.9	3.4
SEN (Action Plus/Statement)	5.4	4.4	4.3	7.8
SEN (School Action)	5.4	4.4	4.3	7.8
Prior attainment (KS2 average points score)	10.0	9.0	8.3	15.9

The missing data, however, are not missing at random, as is evidenced in Table 6.3. More data are missing in the summer and for the separate sciences. Closer inspection of the data reveals that inclusion of the covariate prior attainment completely removes students attending independent schools from the pool; hence the lower percentage of data available for separate science. Furthermore, for English in particular, the attrition rate is greater for modular specifications than for linear. Between 2007 and 2011, 24% of students from independent schools were entered for a single modular specification: OCR's 'GCSE English (Opening Minds) 1900'. This specification was considerably more popular in the independent sector. It attracted only 4% of students from the state secondary sector. Thus, the absence of all students from independent schools also biases the dataset to include a lesser proportion of the modular GCSE results. Findings should be interpreted with this limitation in mind.

Table 6.3 The percentage of data available for analysis in a model of GCSE outcomes controlling for linear/modular assessment route, year-end age, gender, school type, IDACI, a free school meals indicator, SEN indicators and prior attainment

	English				Maths				Science				Separate				
	Linear		Modular		Linear		Modular		Linear		Modular		Linear		Modular		
Year	W	S	W	S	W	S	W	S	W	S	W	S	W	S	W	S	Total
06/07	95.2	89.9	82.3	68.7	93.3	87.0	91.7	93.4	-	77.9	-	93.4	-	53.3	-	91.6	**87.5**
07/08	96.2	89.7	80.3	65.5	93.7	87.8	97.5	93.2	90.0	-	88.5	90.9	-	93.0	-	75.5	**88.2**
08/09	95.8	88.8	85.1	65.5	93.2	87.3	94.5	92.5	-	-	90.9	90.7	-	88.9	97.4	78.3	**88.0**
09/10	95.6	88.3	87.0	63.9	94.0	86.4	94.9	92.4	-	-	86.8	90.7	-	90.7	97.2	81.7	**88.3**
10/11	95.3	87.5	87.3	58.8	94.0	87.3	95.8	91.9	-	-	91.2	90.6	-	-	93.9	83.3	**88.5**
11/12	93.9	79.8	93.0	90.8	93.8	88.0	94.8	89.1	-	88.0	93.8	91.9	-	75.5	77.8	85.8	**90.0**
12/13	92.9	69.9	95.6	91.4	94.3	89.6	95.1	85.4	100.0	54.0	93.8	92.6	89.4	50.6	91.6	89.5	**90.7**
13/14	86.7	89.0	95.3	-	94.3	88.9	92.1	-	80.0	90.1	-	-	94.1	86.9	-	-	**89.3**
Total	**95.5**	**88.9**	**94.0**	**87.0**	**94.0**	**87.8**	**94.8**	**91.8**	**92.6**	**85.7**	**92.9**	**91.6**	**92.1**	**79.5**	**91.6**	**83.8**	**88.8**

Missing data – The best certificated attempt in each academic year (D2)

The formatted dataset includes 4.86 million GCSE English results, 5.20 million GCSE mathematics results, 3.66 million science results and 2.73 million separate science results (see Appendix F, D2.7). It represents a total of 5.14 million students taking examinations in any or all of the subjects of concern. D2 has similar patterns of missing data to those seen in D1. Equivalent tables are included in Appendix G (Tables G1 and G2).

Dependent variables

To measure the effect of linear and modular assessment routes, one outcome of interest is GCSE performance. Another is future performance in an equivalent A level. For the purposes of analysis, GCSE performance is configured in one of two ways.

GCSE grade

For each student the grade is converted to a numeric value and then assumed to be on a continuous scale. Grade A* is given a value of 8, grade A a value of 7 and so on down to grade U which is given a value of 0. In the case of double award qualifications, the intervening grades are given values halfway between the main grades. In other words, grade A*A* has a value of 8, A*A has a value of 7.5, AA a value of 7 and so on.

Grade A to C*

A key grade in school performance measures is the percentage of students who exceed the grade C threshold. This threshold is also often used to determine students' eligibility for further study. A set of binary contrasts have been set up to indicate whether or not a student gained a grade A* to C in each of the subjects of interest. A value of 1 denotes a grade A* to C and a value of 0, grade D to U.

A level grade

For each student for whom there is a matched A level grade, that A level grade is converted to a numeric value and then assumed to be on a continuous scale. Grade A* is given a value of 6, grade A a value of 5 and so on down to grade U which is given a value of 0. The reported outcome measure is mean A level grade in the same subject as the preceding GCSE.

Covariates

Each of the covariates is configured to allow easy interpretation of the results from the modelling. A detailed description of the covariates is given below and the graphs in Appendix H illustrate the GCSE outcome for each level of the covariate separately for modular and linear assessment routes.

Academic year

Academic year is the year in which a student was entered for the given GCSE. The academic year runs from September in the first year until July in the following year. Therefore, an examination taken in November of Year *N* will be grouped with an examination taken in the summer of Year *N+1*. Examinations are nested within academic year but, because there are only eight academic years represented in the dataset and because individual students can take examinations in more than one academic year, academic year is a cross-classifying factor. For English, science and separate science, academic year is almost completely confounded with the linear/modular indicator and is therefore omitted from all modelling.

Series

Series is the time of year at which the examination was taken. In the NPD, this is only differentiated into summer and winter. In some years, the winter series may have included sittings in November, January and March. For modelling, the base category is summer. Figure H1 in Appendix H shows the GCSE outcome for each level of this covariate.

Linear/modular indicator

The linear/modular indicator is a binary contrast set up to identify the assessment route of the GCSE. This is the key covariate of interest in evaluation of the effect of modular and linear assessment routes. For modelling, the base category is the linear assessment route.

Age

Age is the academic year-end age of the student sitting the examination. In other words, it is calculated as age on 31st August in the corresponding academic year. The vast majority of the GCSE students are 16 (Table 6.4), although in science over a quarter of the entry is aged 15 or younger.

Table 6.4 The age distribution of students to which a GCSE certificate was awarded (column percentages)

				First				Best
Age	English	Maths	Science	Separate	English	Maths	Science	Separate
< 16	6.9	12.6	26.4	2.8	12.3	11.9	25.3	2.8
16	92.3	86.6	73.0	96.4	87.0	87.3	74.1	96.4
> 16	0.8	0.8	0.6	0.9	0.8	0.8	0.6	0.9

Age is classified as a factor with three levels: less than16 years old, 16 years old and greater than 16 years old. For modelling, the base category is 16 years old. Figure H2 in Appendix H shows the GCSE outcome for each level of this covariate.

Gender

Because both GCSE English and mathematics are core subjects, the split between male and female results is quite even (Table 6.5). There is some evidence to suggest that male students are more likely to be entered for separate science than female students.

Table 6.5 The gender distribution of students to which a GCSE certificate was awarded (column percentages)

				First				Best
Gender	English	Maths	Science	Separate	English	Maths	Science	Separate
Female	49.8	49.5	50.4	46.7	49.7	49.3	50.3	46.7
Male	50.2	50.5	49.6	53.3	50.3	50.7	49.7	53.3

For modelling, the base category for gender is female. Figure H3 in Appendix H shows the GCSE outcome for each level of this covariate.

School type

School type is coded into two broad categories from the NPD variable KS4_InstitutionTypeCode: state and independent schools. State schools include 6th form and further education colleges and institutions such as special schools, hospital schools and pupil referral units. State schools also include those records where KS4_InstitutionTypeCode is missing. The majority of students are awarded GCSE results having attended a state school (Table 6.6). However, the entry to separate science is slightly more biased towards independent schools. Figure H4 in Appendix H shows the GCSE outcome for each level of this covariate.

Table 6.6 The school type distribution of students to which a GCSE certificate was awarded (column percentages)

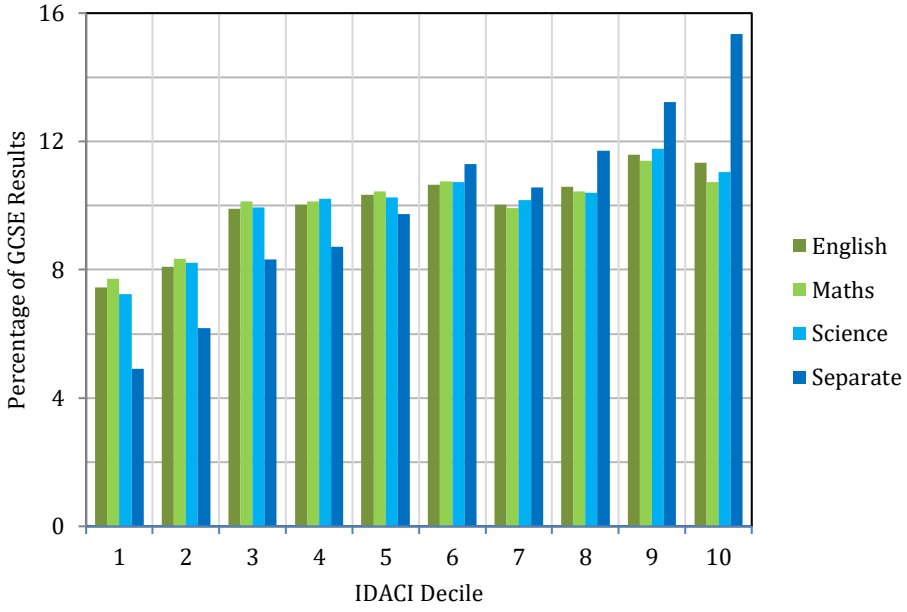
	First				Best			
School Type	**English**	**Maths**	**Science**	**Separate**	**English**	**Maths**	**Science**	**Separate**
State	93.9	95.4	95.6	86.2	94.1	95.8	95.7	86.3
Independent	6.1	4.6	4.4	13.8	5.9	4.2	4.3	13.7

Income Deprivation Affecting Children Index

For given small groups of postcodes,[29] the IDACI measures the proportion of all children aged 0 to 15 living in income deprived families. This index is subsequently split into deciles. The first decile represents those living in the most deprived areas and the tenth decile represents those living in the least deprived areas.

The section of the NPD concerned with examination results and performance tables only includes postcode at a school level. Therefore the IDACI is included as a school-level variable in all modelling of the effect of assessment route. As such it is a proxy for the deprivation of the students attending the school. There will naturally be cases where schools situated in deprived areas attract privileged students and vice versa. This should be regarded as a limitation of the data.

The IDACI is centred so that the baseline category is six, representing a school in the middle of the range in terms of deprivation. For the GCSE results included in the analysis the distribution across each of the deciles is slightly biased towards the less deprived areas (Figure 6.3). Figure H5 in Appendix H shows the GCSE outcome for each value of the IDACI covariate.



[29] Lower Layer Super Output Areas (LSOA)

Figure 6.3 The IDACI distribution of students to which a GCSE certificate was awarded (D1)

Free school meals, SEN (Action Plus/Statement) & SEN (School Action)

Free school meals, SEN (Action Plus/Statement) & SEN (School Action) are all binary contrasts which identify that a student has been in receipt of additional support. A student will be given free school meals if they or their guardian is eligible for income support or other such benefits. This variable is often used as a proxy for socio-economic status. For modelling, the base category for free school meals is a child *not* in receipt of this benefit.

SEN (Action Plus/Statement) & SEN (School Action) identify students who have special educational needs. SEN (School Action) indicates that a student with SEN has been provided with interventions that are additional to, or different from, those provided as part of a school's usual differentiated curriculum offer and strategies. SEN (Action Plus/Statement) indicates that a student with SEN has been provided with interventions by the school with the support of external agencies and specialists (DfES, 2001).[30] For both of these variables the base category is a student *without* any special educational needs. Figures H6, H7 and H8 in Appendix H show the GCSE outcome for each level of these binary contrasts.

Prior attainment

Student prior attainment is measured as the average Key Stage 2 fine points score. In the dataset there were a number of records with a prior attainment of zero. This is a highly unlikely score. A student scoring a single mark in each of the SATs papers and 'working towards level 1' in writing would have a prior attainment of 11. Therefore the prior attainment for students with a score of zero was set as missing. A small number of students were assigned scores greater than the maximum of 40. The prior attainment for these students was also set to missing. For D1, Figure 6.4 shows that the distribution of prior attainment scores across the academic years of interest was broadly similar but that there was slight skew towards the top end of the points range. Figure H9 in Appendix H illustrates the GCSE outcomes for all values of the covariate but, for modelling, the prior attainment measure has been normalised.

[30] These terms have been discontinued since the 2015 Special Educational Needs and Disability Code of Practice.

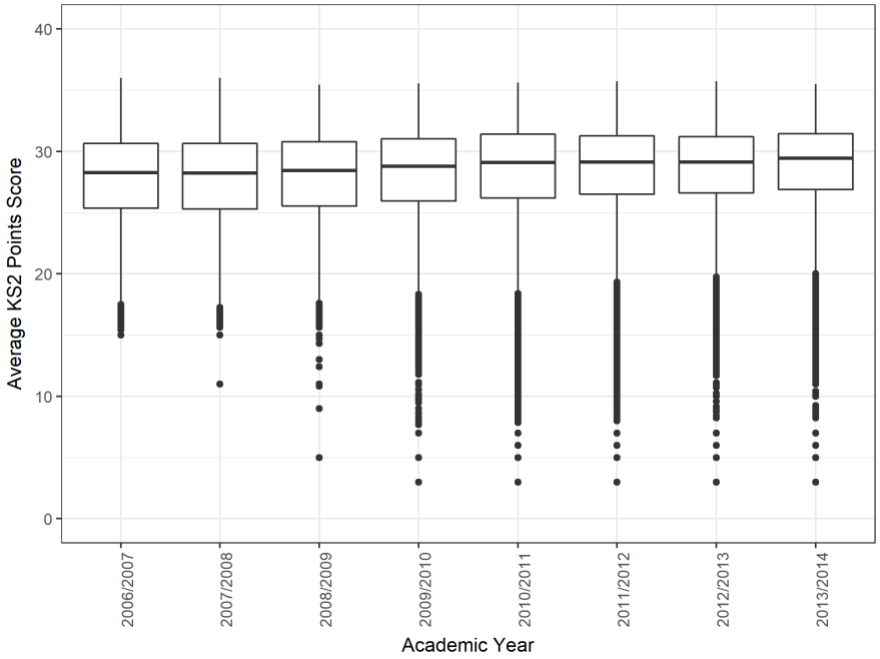


Figure 6.4 The distribution of average Key Stage 2 points scores by academic year (D1)

Multicollinearity

Multicollinearity can affect estimation of the relationship between covariates and a dependent variable. To this end, Figure 6.5 shows the correlation between covariates chosen to model the assessment route; the higher the correlation between two variables, the larger the dot. A negative correlation is green in colour and a positive correlation, blue. The correlation is relatively low for all variables.

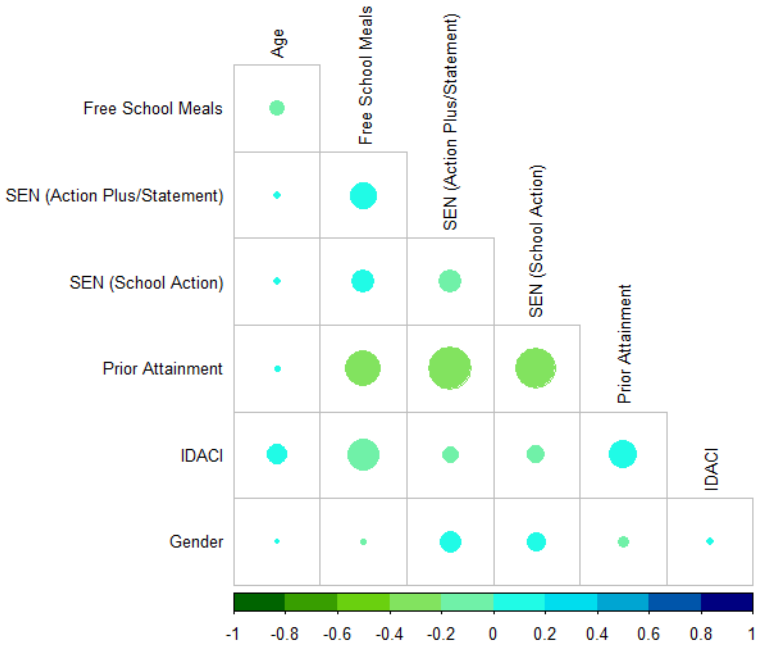


Figure 6.5 The correlation between covariates (D1)

Assessment route and school type performance

Historically, Key Stage 2 performance has been shown to explain a significant proportion of variation in GCSE outcomes (see, for example, Strand, 2006). However, no prior attainment data are routinely collected for students from independent schools. This prevents robust modelling of their performance in linear and modular examinations. As such, a simple comparison of the GCSE outcomes for state versus independent schools is presented (using D1.7 and D2.7 as described in Appendix F).

To reduce bias brought about by non-random entry to assessment route, the data have been matched to create quasi-experimental conditions. The matching allows fairer comparisons of the effect of the linear and modular assessment routes on GCSE outcome.

Because all of the available matching covariates are discrete, frequency matching[31] is used to control for gender, subject, age group, school type and IDACI; making 480 distinct categories for comparison. Each of these categories is populated by at least one student entered for a linear examination and at least one for a modular examination. Over 95% of the categories have more than 20 students following each assessment route.

Within each category, the difference in mean outcome for modular and linear students is calculated. The category means are weighted to reflect the prevalence of the category within the population. A final mean difference in the performance of modular and linear students is given by summing the weighted category means.

Mean GCSE result

Figure 6.6 illustrates the difference in mean grade outcome for modular and linear assessment routes. A positive value denotes that, all other controlled covariates being equal, the modular outcome was better. On the other hand, a negative value denotes that the linear outcome was better. The bars with solid fill represent the comparison for the first certificated attempt at the GCSE in question (D1). The shaded bars represent the best certificated attempt in each year (D2). The blue bars show the picture for state schools and the purple bars for independent schools.

[31] This is sometimes termed category matching (Anderson et al., 1980) or sub-classification matching (Stuart, 2010). When the categories are broad, discrete and completely populated, individual matching is unnecessary and leads to a loss in statistical efficiency. Frequency matching has been shown to reduce bias in estimates of the difference between treatment and control groups (Cochran & Rubin, 1973).

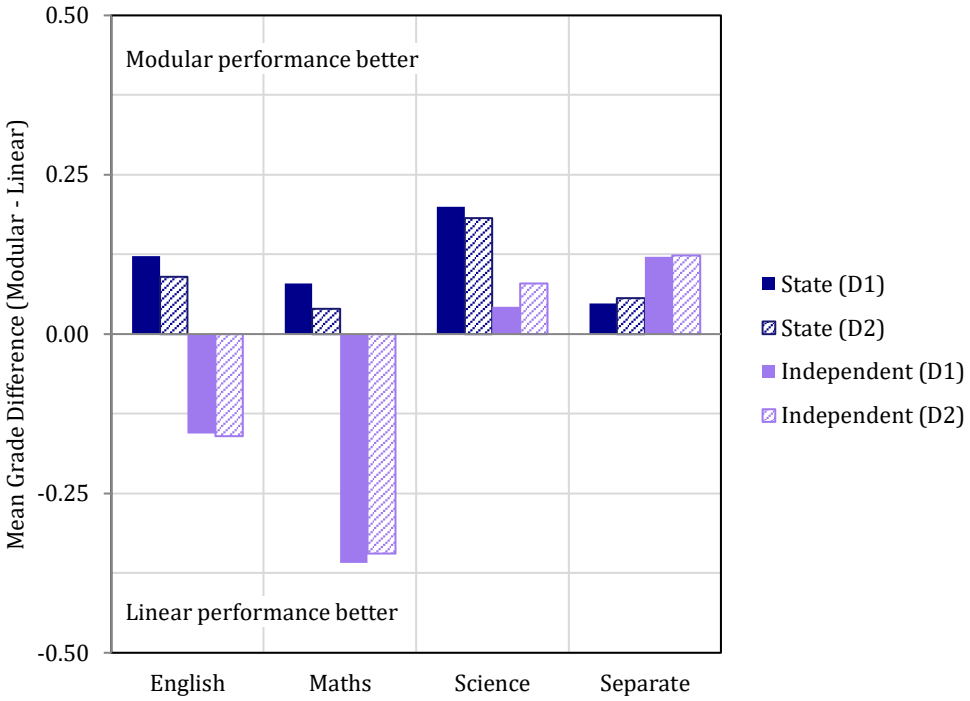


Figure 6.6 The mean difference in GCSE grade between the modular and linear routes dependent on school type of entry

The first point of interest is that, for no subject is the difference in GCSE outcomes between the modular and linear assessment routes greater than $^{3}/_{8}$ of a grade. For most it is nearer to $^{1}/_{8}$ of a grade.

Within state schools, the students appear to perform better in modular assessments than in the linear equivalent. The pattern is more pronounced when only the first certificated attempt at the GCSE is considered (D1). However, hidden within the first certificated attempt for a modular GCSE is the opportunity for unit re-sits. Once the possibility of linear recertification is included in the analysis (D2), the difference between modular and linear outcomes is greatly reduced.

Within independent schools, the pattern is very different. For all but science, the mean grade differences are similar regardless of whether the first attempt (D1) or the best attempt in a year is considered (D2). This is likely to be due to differences in recertification patterns between the state and independent sector. Across all the subjects under consideration, the proportion of state school students who opt to re-sit (with certification) is 12.0% whereas the same statistic for independent school students is 1.5%.

Furthermore, in English and mathematics the independent school students appear to perform better following a linear assessment route than a modular assessment route. The pattern in science and separate science is more akin to that seen in the state sector.

Percentage of A to C awards*

Figure 6.7 illustrates the difference in percentage of grade A* to C awards for modular and linear assessment routes. The patterns are very similar to those seen in Figure 6.6. The difference between the modular and linear assessment routes is never greater than 7% and, in the separate sciences, is less than 1.5%. On the whole, state school students appear to perform better in modular examinations. Independent school students tend to do better in linear English and mathematics GCSEs.

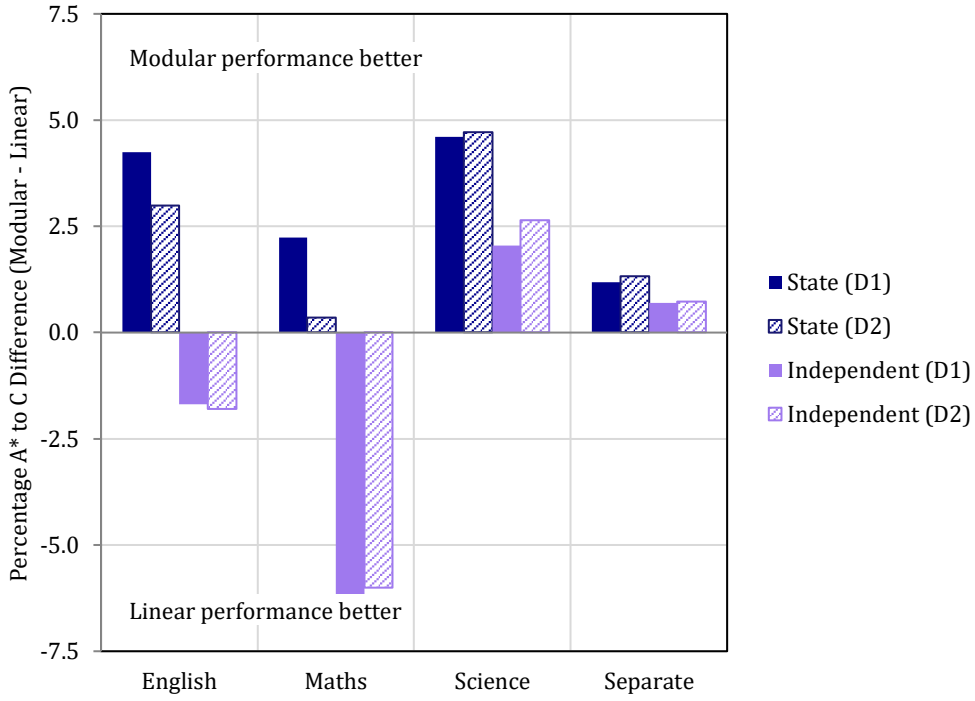


Figure 6.7 The mean difference in percentage of grade A*to C awards between the modular and linear routes dependent on school type of entry

Discussion

While these findings provide an interesting view of the performance of state and independent sector schools, the comparisons do not account for the single, most effective covariate: prior attainment. Whenever a matching technique is used, the central assumption is that of strongly ignorable treatment assignment (Rosenbaum & Rubin, 1983). With regard to the comparison of linear and modular assessment routes, this means that there is no relationship between unobserved covariates and the allocation of students to GCSE specifications. With the effect of prior attainment remaining uncontrolled, it cannot be said there is strongly ignorable treatment assignment. It is quite possible that schools choose different assessment routes for their students based on *a priori* knowledge of their ability. This is compounded by the fact that students are selected to attend independent schools on the basis of their prior attainment.

Returning to the unmatched data, it is clear to see that there are differences in the proportion of linear entries between school types. For example, the independent school entry to linear GCSE English is relatively low compared with the state school entry (Table 6.7). This, as alluded to earlier, is likely to be due to

a preference for OCR's modular GCSE English (Opening Minds) among independent schools. Any systematic, unmeasurable difference between assessment routes undermines the validity of the analysis. Therefore, it would be wrong to generalise from Figure 6.6 and Figure 6.7 that English and mathematics students from the independent sector will perform better in linear examinations. It would mean ignoring the possibility that weaker students are entered for modular examinations.

That said, it is noteworthy that independent schools enter students for fewer re-sits (with certification) and, within these schools, the pattern of relationship between modular and linear outcomes is not consistent across all subjects.

Table 6.7 The percentage of entries made to linear specifications between the academic years 2006/2007 and 2013/2014

School type	English	Maths	Science	Separate
State	68.9	64.2	10.5	18.2
Independent	57.5	71.9	23.2	23.4

Assessment route, the gender gap and the effect of socio-economic status

The interaction between prior attainment, choice of assessment route and GCSE outcome can be explored in greater detail for schools where Key Stage 2 testing is compulsory (using D1.8 and D2.8 as described in Appendix F). Models can be used to control for factors, described in the 'covariates' section above, such as gender, socio-economic status and examination series.

Limitations of the data

Because the data on the NPD are observational in nature, there is no natural balance between the linear and modular entries; the data cannot be said, even closely, to represent a randomised control trial. Raw differences between modular and linear assessment routes could, therefore, be distorted as a result of selection biases in entry patterns. Before analysis, it would be preferable to create quasi-experimental conditions; matching the data such that, for each assessment route, the data are comparable on measured covariates that have the potential to influence GCSE outcome (Ho et al., 2007). However, computational limitations arising from the sheer volume of data on the NPD preclude the use of matching for every potential model.[32]

In order to understand the effect of any potential biases, therefore, a test comparison between the findings using matched and unmatched data has been performed using the separate science data (D1). The separate sciences have been chosen for two reasons. Firstly, the entry is smaller and therefore the dataset is more manageable. Secondly, and more importantly, because separate science is

[32] For each of the four subjects, there are two datasets (D1 & D2) with a minimum of 2 million records and two possible dependent variables.

likely to be the subject that is most influenced by selection bias. The choice to take the separate sciences is often ability- or opportunity-based and may also be regarded as influenced by aspiration (Alldritt & Taylor, 2014).

Creating a matched dataset

The variable over which matching is performed is assessment route; the linear/modular indicator. The covariates are age group, examination series, gender, IDACI, free school meals, SEN (Action Plus/Statement), SEN (School Action) and prior attainment (KS2 average points score). Because of the size of the dataset, full matching is used to produce a near-optimal solution in a fraction of the time needed to create a nearest neighbour match (Sävje et al, 2016; Sävje et al, 2017). Each record is assigned a weight to effect a balance between covariates on the matching variable and, thus, this method of matching has the advantage that all records are retained (Stuart, 2010).

It should be noted that assessment route is almost entirely confounded with school of entry. In other words, schools generally choose one specification, which is either linear or modular, for all of their students to follow. Therefore, the matched data fail to reflect the hierarchy inherent in the structure. Nevertheless, the hierarchy can be controlled in subsequently fitted multilevel models; so too can the effect of any potential interactions between the matching variable and other covariates.

Comparison of matched and unmatched models

In the unmatched data, the largest normalised difference between the linear and modular cohorts of students is in the age profile and the smallest is in the measure of prior attainment. The differences are all reduced by the matching process. In fact they are almost eliminated for all covariates except the series of entry: summer or winter. The effect of matching the data is illustrated in Figure 6.8.

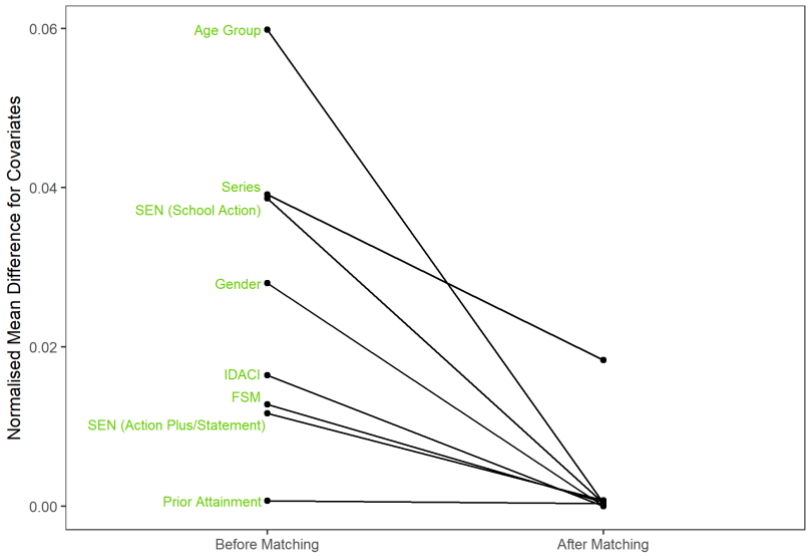


Figure 6.8 The normalised difference of means of each covariate before and after matching (separate science D1)

A comparison of the parameter estimates for a matched and unmatched model of GCSE grade outcome allows an evaluation of the extent to which matching reduces bias. Details of the model specification are described in the next section. Figure 6.9 shows the parameter estimates derived from the matched and unmatched data with error bars at 1.96 times the corresponding standard error.

For all covariates but the series of entry, the parameter estimates are very similar and the error bars overlap. This suggests that, in terms of interpreting the effect of assessment route on GCSE outcome, there is not much to be gained by using matched data. The difference in the parameter estimate for the covariate series is largely due to the very small proportion of students first certificating in separate science in the winter series (0.2%). The number of winter certifications is higher in the other three subjects: 1.9% for combined science, 13.7% for English and 20.4% for mathematics. Therefore any benefits of matching are likely to be even smaller.

Overall, it seems unlikely that substantive conclusions would be affected by the use of unmatched data in the modelling. Therefore, due to the computational limitations arising from the size of the datasets, unmatched data are used for all analysis of the effect of assessment route on GCSE outcome.

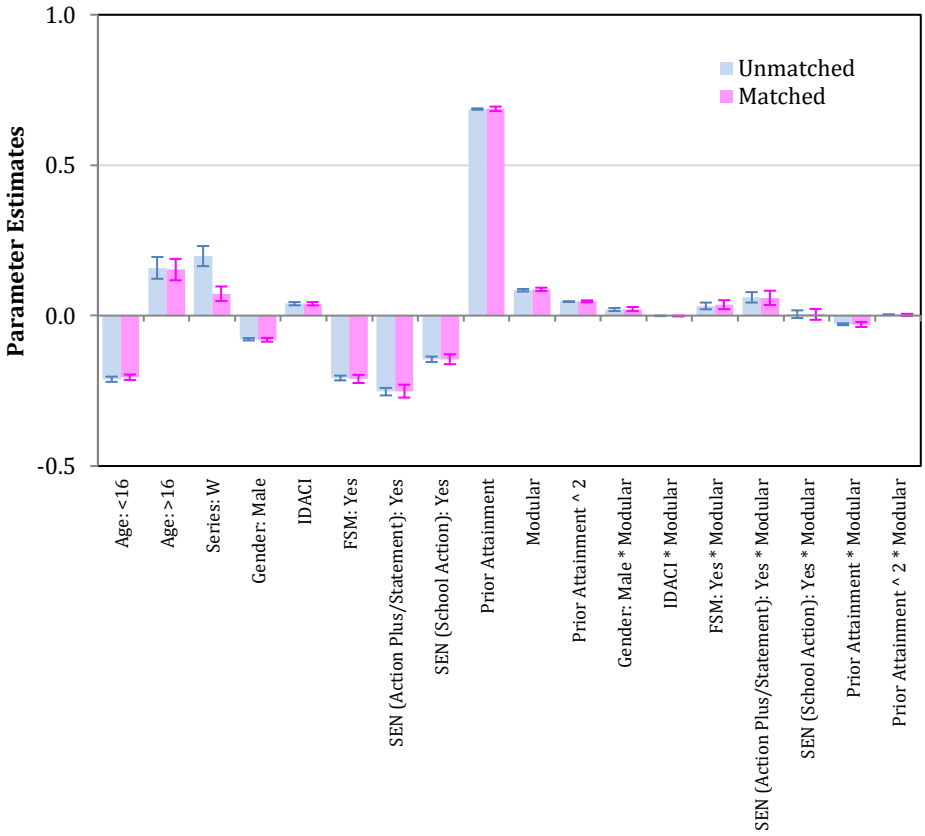


Figure 6.9 A comparison of the parameter estimates derived from a model of matched and unmatched data (Separate science D1)

Multilevel models

For each of the subjects and each of the datasets (D1 & D2), two multilevel models are fitted to evaluate the impact of assessment route on GCSE outcome.[33] The first is a linear model which has GCSE grade as a continuous dependent variable (Box 1).

$$GCSE\ grade_{ij} = ß_{0j} + ß_1 x_{ij} + \ldots + ß_n x_{ij} + e_{ij}$$
$$ß_{0j} = ß_0 + u_{0j}$$
$$u_{0j} \sim N(0, \sigma^2_{u0})$$
$$e_{0ij} \sim N(0, \sigma^2_e)$$

Box 1

The second is a logistic model of the probability of being awarded a grade A* to C (Box 2).

$$log\left(\frac{Prob.A*C_{ij}}{1\text{-}Prob.A*C_{ij}}\right) = \beta_{0j} + \beta_{1j} x_{ij} + \ldots + \beta_{nj} x_{ij} + u_j$$
$$ß_{0j} = ß_0 + u_{0j}$$
$$u_{0j} \sim N(0, \sigma^2_{u0})$$

Box 2

In both models, GCSE outcome is nested within school of entry. In D1 there is only one GCSE outcome per student so the level one unit is the student. In D2, each student could appear more than once so the level one unit is the certificated GCSE outcome.

The models are fitted using the lme4 package for R (Bates et al, 2015). For each model, all educationally relevant covariates have been retained regardless of significance. In addition to the main effects, the models include a quadratic term for prior attainment and interactions between the main effects and the assessment route. Tables including details of parameter estimates, along with estimates of variance explained (Snijders & Bosker, 1994) and the variance partition coefficient[34] (Goldstein et al, 2002), are given in Appendix I (Tables I1 to I4).

A note about standard errors, significance and effect sizes

Inevitably, when dealing with GCSE population data in EBacc subjects, the volume of data means that even small differences in outcome between subgroups of students will, in some sense, be statistically significant. The definition of this statistical significance and any subsequent inference of educational significance, however, require some unpacking.

With the introduction of complex modelling, the wish to summarise data by means of p-values has given much cause for debate (Bolker, 2017). The lme4

[33] For details of how GCSE outcome is configured see the 'Dependent variables' section.
[34] Often referred to as intra-unit correlation.

package for R does not provide p-values for the fixed effects. This is because, in multilevel models with unbalanced data, the ratio of the parameter estimate to the standard error has an unknown distribution (Bates et al., 2015). Although this ratio can be calculated, it cannot be said to follow a t-distribution. Furthermore, and specifically with reference to the logistic regression (Box 2), the chosen optimisation routine for the models means that evaluation of the standard errors is approximate.[35] Therefore, instead of reporting p-values in the models discussed below and tabulated in Appendix I, a † has been placed beside each covariate where the ratio of the parameter estimate to its reported standard error is greater than two. This approximate approach to statistical significance is only really appropriate given the volume of data involved.

Given the limited evidence provided by any significance testing, the effect size of each covariate is also reported. Effect size is used to determine the relative importance of a covariate in explaining variance in GCSE outcome. Despite the well-established name, the effect size does not necessarily imply a causal link; it merely allows valid comparison of the magnitude of each effect or covariate. Indeed, Coe (2004) described an effect size as 'a standardised scale-free measure of the relative size of the effect of an intervention'. He also advocated that effect size should be reported with confidence intervals. While confidence intervals have been calculated for all effect sizes associated with the models of GCSE outcome, they are so small as to be negligible and have therefore been excluded from the reporting.

Effect sizes for the linear model have been calculated following the method outlined by Tymms, Merrell and Henderson (1997) and expanded upon in Strand (2004) and Tymms (2004). They are defined as the difference in mean score expressed in standard deviation units.[36] Effect sizes are often reported as small (0.2), medium (0.5) or large (0.8) with reference to Cohen's original work in this area (Cohen, 1969). These broad classifications are not appropriate in the models of GCSE outcome because the large volume of data means 'small' effect sizes could still have a widespread impact on students. The effect sizes are simply reported to clarify the relative importance of each covariate.

Effect sizes for the logistic model are reported as odds ratios (Rosenthal, 1996). The odds ratio is not always a standardised measure because it depends on the configuration of the covariate as to how it should be interpreted. As all of the covariates in the analysis of GCSE outcome are either binary or standardised;[35] direct comparison between quoted odds ratios is valid in this analysis.

[35] The argument `control=glmerControl(optimizer = "nloptwrap", calc.derivs = FALSE)` is used to turn off the time consuming derivative calculation that is performed after the optimisation is finished. This speeds up the processing time considerably but disables some of the convergence tests, as well as making lme4 use a less accurate approximation to compute the standard errors of the fixed effects. See https://cran.r-project.org/web/packages/lme4/vignettes/lmerperf.html.

[36] IDACI is not standardised; it is merely centred, so an adjustment has been made to the effect size such that it is comparable with the effect sizes for the other covariates (see equation 11 in Tymms (2004)). A similar adjustment has been made to the odds ratio calculations for the logistic model.

Mean GCSE result

Effect sizes for each of the linear models of GCSE outcome are presented graphically in Figure 6.10. It is clear that, no matter what the subject and no matter whether recertification is considered, the main factor associated with success in the EBacc GCSE subjects is prior attainment. The strongest relationship is in mathematics and this is reflected in the high R^2 for the model (0.657 for D1 and 0.649 for D2).

A consistent finding across all subjects is that being in receipt of free school meals or having some form of special educational needs is associated with a poorer GCSE outcome. Furthermore, the effect size for IDACI suggests that if a school is positioned in a more disadvantaged area, the students will gain lower GCSE grades.

Interestingly, for English alone, gender seems to have a relatively high association with success. Even after accounting for prior attainment, female students perform better in English than their male counterparts.

Any effect of assessment route is almost non-existent for all but separate science. Even with separate science, the effect size is relatively small compared with that for other covariates. To a certain extent it is also confounded with the limited availability of linear assessment (Figure 6.1 and Figure 6.2) over the period of study.

Interaction effects, although largely statistically significant, add little information about the factors influencing GCSE outcome.

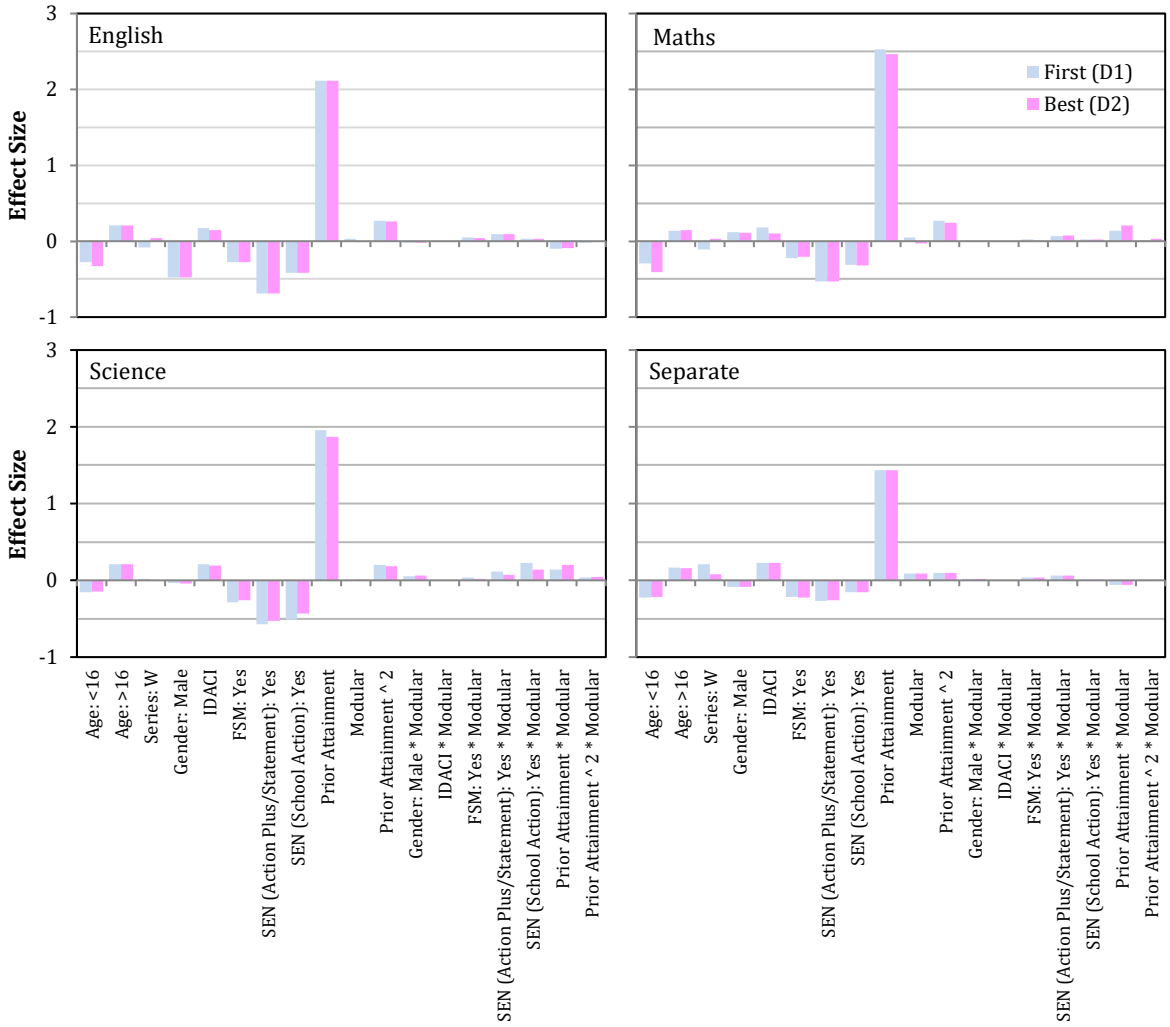


Figure 6.10 Effect sizes for the multilevel linear models of GCSE outcome

Percentage of A to C awards*

Odds ratios for each of the logistic models of the probability of gaining a grade A* to C are presented graphically in Figure 6.11. Broadly, the findings are the same as for the linear model. The fit of each model is similar to its linear equivalent. Prior attainment appears to dominate future GCSE prospects. The influence of other covariates is minimal in comparison.

One feature of the logistic models is that the variance partition component is higher than that seen for the linear model. This statistic is reported for the null model and describes the percentage of variation in outcomes that is attributable to the school before any other factors have been accounted for (Appendix I). School, it would seem, has a higher influence on whether a student will obtain a grade A* to C than it does on the actual grade a student is awarded. While this could be related to a school's approach to accountability measures, it is more likely to be due to the dichotomisation of the grade range in the logistic model.

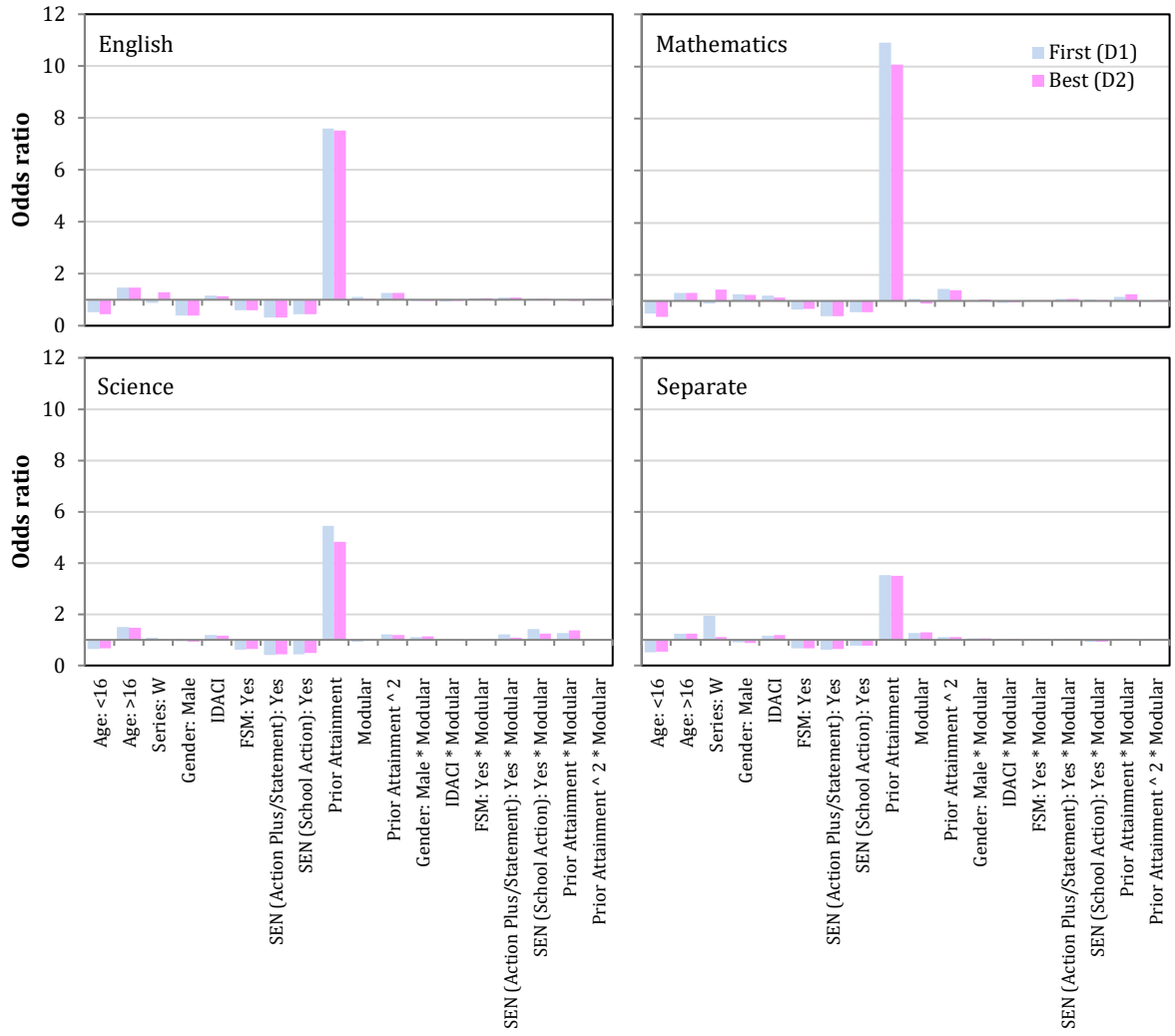


Figure 6.11 Effect sizes for the multilevel logistic models of the probability of gaining a grade A*to C

Discussion

On linear and modular teaching and assessment

The intention of formulating two separate datasets to analyse the impact of a linear or modular GCSE assessment route was to consider aspects of both teaching and testing. The first dataset was designed to investigate the effect of teaching a modular or linear course up until the point of first certification (D1). The second was designed to allow exploration of the effect of assessment route combined with recertification opportunities. The two datasets give very similar findings but, on the whole, the effect sizes are smaller for D2 where recertification opportunities have been exploited.

In the linear model, the main effect parameter estimates for modular assessment (compared with the baseline of linear assessment) are as shown in Table 6.8. Clearly after recertification opportunities are exploited, there is almost no difference in the grades awarded to GCSE English and combined science students dependent on assessment route. The difference is a thirtieth of a grade for the

mathematics students (-0.034) and just under a tenth of a grade for separate science students (0.087). As mentioned earlier, however, the findings for separate science are confounded with the limited availability of linear assessment over the period of study.

Table 6.8 Main effect parameter estimates for modular assessment (compared with the baseline of linear assessment)

	English		Maths		Science		Separate	
	Estimate	**(se)**	**Estimate**	**(se)**	**Estimate**	**(se)**	**Estimate**	**(se)**
D1	0.031	(0.002)	0.057	(0.002)	-0.003	(0.003)	0.084	(0.003)
D2	-0.001	(0.002)	-0.034	(0.002)	0.001	(0.003)	0.087	(0.003)

The interaction effects between assessment route and other covariates are all small but perhaps the most notable is the interaction with prior attainment. Figure 6.12 shows the pattern of relationship between prior attainment score and GCSE grade outcome for the best GCSE outcome in a given academic year (D2). This pattern differs subtly between subjects but, nevertheless, suggests almost no educational advantage can be gained by entering students to modular or linear assessment routes based on prior attainment.

For mathematics and combined science there is limited evidence to suggest that a modular assessment route might favour students with higher levels of prior attainment. Conversely, a linear assessment route might favour students with a lower level of prior attainment.

Modular and linear assessment routes, however, are inextricably intertwined with unmeasurable choices that are made in the course of a student's education. GCSE specifications perceived to have high utility for further study might be selected by a school regardless of the teaching and assessment structure.[37] Modular specifications may be taught in a linear manner and vice versa. Linear examinations might be offered to poorer students to maximise teaching time; alternatively the same students might be offered modular assessment to reduce the burden of assessment. Choices might not be choices; schools might be limited by resources or by the availability of specifications. After all, there were no linear combined science specifications available between 2007/2008 and 2012/2013. Furthermore, hidden beneath each modular certification are many unit re-sit opportunities. All of these factors, and more, make it difficult to interpret any differences in modular and linear assessment routes.

[37] This issue is discussed in more detail in the next section: Assessment route and the effect on A level uptake and outcome.

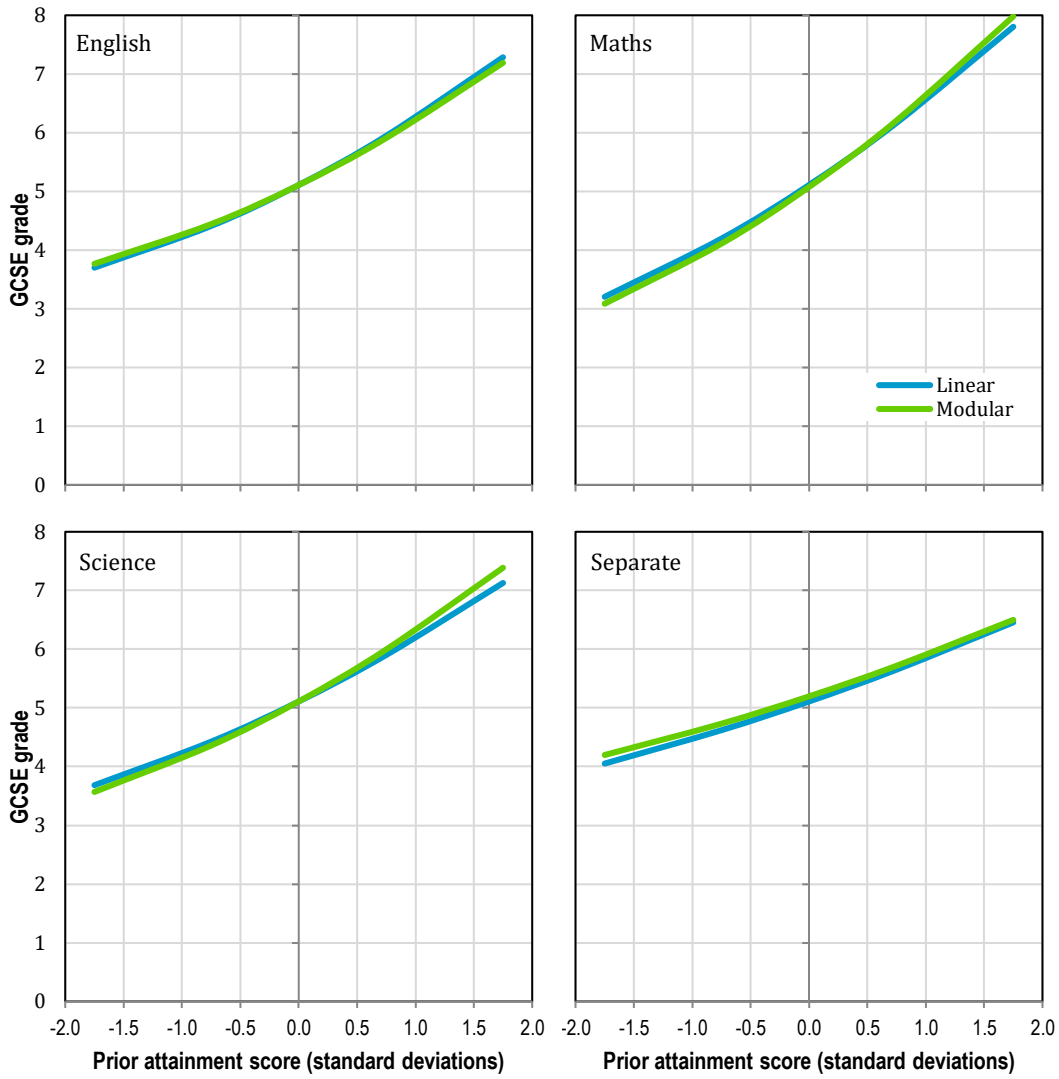


Figure 6.12 The relationship between prior attainment score and GCSE grade outcome (D2)

On the gender gap and effect of socio-economic status

The effects of gender and socio-economic status on GCSE performance are small in comparison with the effect of prior attainment. However, students' formative experiences are all hidden within the measure of prior attainment. While the effect of gender on GCSE outcome appears small, it may only be small because the measure of prior attainment conceals the same biases. Effect sizes, therefore, reveal only patterns that have emerged in the years between Key Stage 2 assessment and GCSE assessment.

Thus, the comparatively large gender effect for GCSE English represents a widening of the gap between male and female students over the course of their secondary education. Even after accounting for recertification opportunities, female students have almost a half grade advantage over their male counterparts. The interactions effects between gender and assessment route are sufficiently small to suggest no real advantage, or otherwise, of entering students to specifications based on gender.

The models suggest that social disadvantage is associated with poorer performance in GCSE assessment. They do not, however, support the notion that a particular assessment route favours the disadvantaged. While the interactions are significant in a statistical sense, they are unlikely to be of any real educational significance.

Assessment route and the effect on A level uptake and outcome

The NPD data available for evaluation of the effect of modular and linear assessment include Key Stage 5 information from 2007 until 2015. Therefore, assuming a traditional two-year route to certification, A level data can be matched for all but the 2013/2014 GCSE students. The number of matched students is recorded on the flow diagram in Appendix F. Analysis of A level performance refers only to a student's best certificated GCSE attempt in each academic year (D2, D2.10 in Appendix F), the reason being that best GCSE outcome will usually provide the gateway to A level study.

A level uptake

The proportion of GCSE English, mathematics and science students continuing to A level in the same subject area differs between subjects (see row 1 of Table 6.9). Over a third of the students entered for GCSE separate sciences are subsequently awarded an A level grade in science. Perhaps because English and mathematics are compulsory in the national curriculum, uptake of the comparable A level is much lower. Only about 10% of GCSE English students go on to take A level English and even fewer GCSE mathematics students continue to A level mathematics. It is the GCSE science students, following a combined science route, who are least likely to enter an A level in science. The probable reason is that the combined science GCSE qualifications tend to attract a weaker cohort of science students.

The uptake of A level also seems to differ slightly dependent upon GCSE assessment route (see rows 2 and 3 of Table 6.9). Linear GCSE English students are slightly more likely to be awarded an A level English grade while, in GCSE mathematics, it is the modular students who are more likely to certificate. For the science subjects it is the linear route from which students are most likely to proceed to A level. However, it is worth noting that in science there is only one academic year – 2006/2007 – where linear GCSE examinations were widely offered and where there was matched A level data available.

Table 6.9 The percentage of students continuing from GCSE to A level in a comparable subject

	English	Maths	Science	Separate
All GCSEs	11.6	8.4	6.2	38.6
Linear GCSEs	11.9	7.8	8.7	45.5
Modular GCSEs	10.9	9.2	6.1	38.4

Many factors contribute towards the decision to take an A level and some might even be intertwined with the previous choice of GCSE specification. What is interesting, though, is that there is very little difference in the mean GCSE result for the linear and modular students opting for A level (Table 6.10). This seems to imply that, where students have comparable GCSE grades, it is not the assessment route which determines the decision to continue to A level.

Table 6.10 Mean GCSE result for students proceeding to A level in a comparable subject*

	English	Maths	Science	Separate
Linear	6.5	7.3	7.0	7.0
Modular	6.5	7.3	6.8	7.0

* Despite being small, differences in mean GCSE result between linear and modular assessment routes are statistically significant because of the large number of students.

It would appear, therefore, that assessment route has the potential to influence future prospects only where there are concerns of differences in the probability of gaining an *A level enabling* GCSE grade.

A level attainment

One constant throughout the students' GCSE teaching is their measure of prior attainment. Over 95% of students proceeding to A level in a comparable subject were awarded an average Key Stage 2 score between 25 and 35. Figure 6.13 shows the mean A level grade by Key Stage 2 performance dependent upon assessment route of the GCSE. The shaded regions, where large enough to be visible, indicate the 95% confidence bands. Although there are some small differences in A level outcome, most of these differences amount to less than a tenth of a grade. So, although statistically significant because of the number of students involved, they are likely to be of minimal educational significance. Furthermore, where the differences are at their largest, the modular and linear cohorts are most disparate in terms of entry numbers.

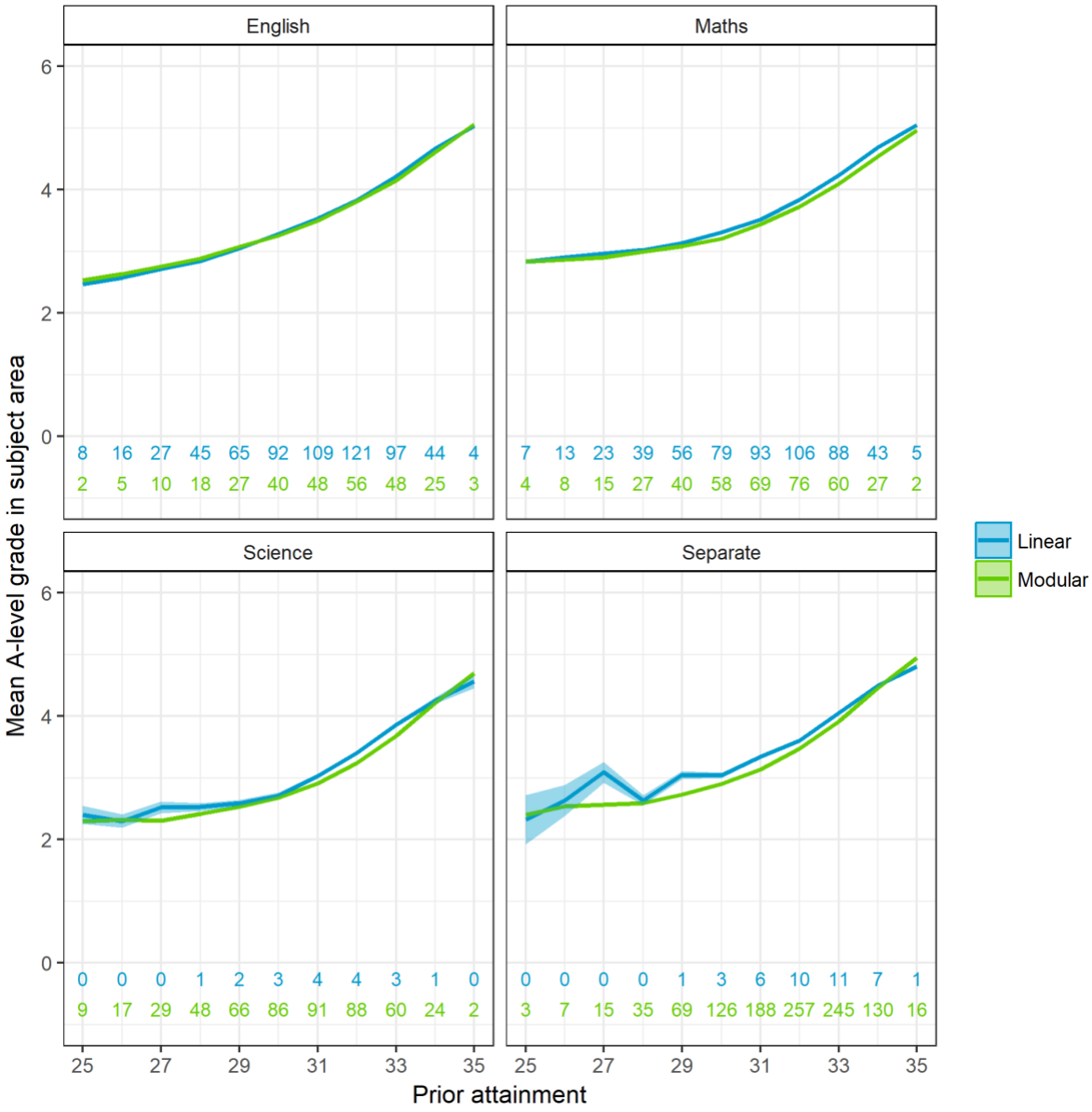


Figure 6.13 Mean A level grade by prior attainment with shaded confidence bands. Data in table denotes the number of entries in thousands for linear and modular assessment routes (D2)

Discussion

The extent to which GCSE assessment route influences the outcome at A level appears minimal. More significant perhaps is the extent to which different GCSEs enable, or are perceived to enable, progression to A level. Although it is tempting to attribute the differing progression rates in Table 6.9 to assessment route, these differences may arise at the more granular level of specification choice. Figure 6.14 shows the ten highest entry GCSE English specifications (2006/2007 to 2012/2013) overlaid with the percentage of students proceeding to A level. There are distinct differences in A level uptake between the specifications and these differences seem independent of the assessment route.

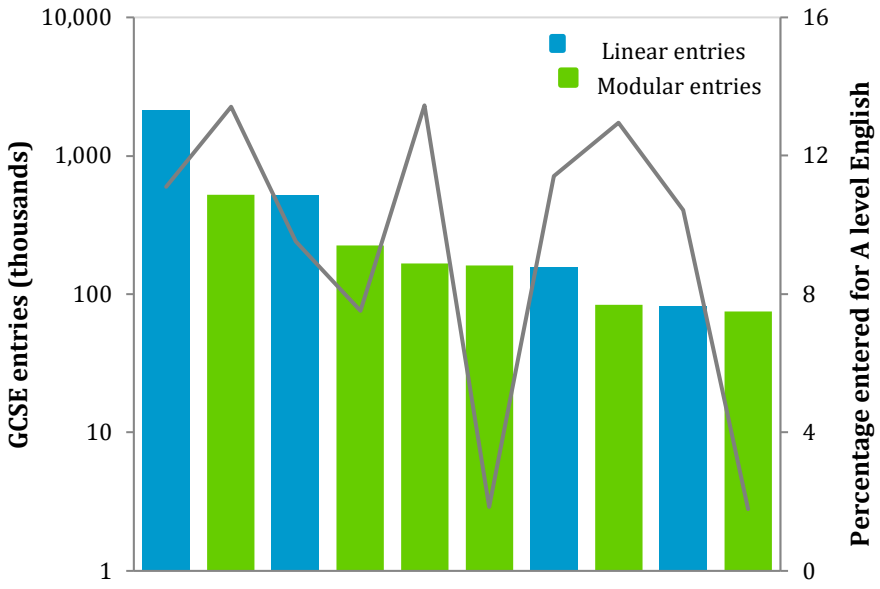


Figure 6.14 GCSE English entries and percentage progression to A level for the top 10 most popular GCSE specifications 2006/2007-2012/2013 (D2)

So, as far as further study is concerned, it would appear that the content and structure of the individual GCSE specification is likely to have greater influence on progression to A level than modularity or linearity.

Summary of pupil level quantitative analyses findings

The quantitative evaluation of the effect of assessment route on GCSE outcomes is limited to the extent that data are not collected under experimental conditions, nor could they ever be. When considering the conclusions that are reached, there are a number of important caveats.

- Modular and linear assessment routes were not available in each subject in every academic year. Therefore, the outcomes for modular and linear assessment might be confounded with the standard setting process.
- Modular examinations allow unit re-sits before certification. This is not captured in the NPD and therefore the analysis reflects GCSE outcomes to a greater extent than the effect of modular and linear teaching.
- Missing values in the data are not missing at random.
- Covariates for socio-economic status are all proxy measures.
- There are many unmeasurable covariates which may affect GCSE performance. Not only might teacher, school choices and specification availability impact outcome, so too might more subtle student-level influences.
- The models are compromised because of the computational limitations of dealing with population level data.

With these caveats in mind, the evidence suggests that there is almost no difference in the performance of students who follow modular and linear routes in GCSE English, mathematics, combined science and separate science.

- In GCSE English there is some evidence that the gap between male and female students widens during secondary education. Female students perform better than their male counterparts. Patterns are far less pronounced in mathematics, combined science and separate science. There is no educationally significant evidence that assessment route influences male and female students differently.
- The models suggest that social disadvantage is associated with poorer performance in GCSE assessment. There is no support, however, for the notion that a particular assessment route favours the disadvantaged.
- Differences in the performance of state and independent schools are based on a less robust model. The model implies that state school students perform better in modular than linear assessments. The picture is less clear cut for independent schools. However, given the disparity in specification entry patterns between state and independent schools, it is possible that the effect of school type is confounded with specification choice rather than modularity or linearity.
- GCSE assessment route has little effect on A level attainment in the same subject. It is much more likely that the content and structure of the individual GCSE specification has an influence on progression to A level, than its modular or linear nature.

Given the evidence available from the NPD, it is unlikely that students' GCSE results have been unduly influenced by the assessment route. Next, we turn to research with teachers on their beliefs about and reactions towards the examination structures.

Chapter 7 Baseline research in schools

The qualitative research strand of the project investigated the following research questions:

1. What do teachers believe the effects of examination route are upon overall GCSE outcomes and grade inflation?
2. What do teachers believe the effects of examination route are for different groups of pupils?
3. How, if at all, has the examination route influenced teaching strategies and teaching materials according to the teachers?

Two waves of interviews with teachers were planned. Since the availability of modular GCSE examinations had decreased, the first wave was conducted as early as possible in the project, and the second one once the new linear specifications had been introduced (2016-17). Ideally, the same schools would be revisited in the second wave to investigate whether there had been any changes in practice. The first wave of interviews were intended to collect data on teachers' retrospective views of their teaching and administrative behaviour in relation to the previously modular specifications and any initial changes that had occurred since the specifications had been made linear.

Method

Sampling

In the first wave of the data collection participating schools were selected using a random stratified sample divided equally between schools in England who entered students for modular and linear qualifications in mathematics in 2013 (the last year in which modular qualifications could be taken in a modular manner). Sampling was stratified by school type and by the kind of GCSE examination entry policy institutions appeared to have when modular GCSEs were available. If schools entered 50 per cent or more of their GCSE mathematics candidates for a modular qualification in 2013, they were considered to have a mainly modular entry policy. If less than 50 per cent of the pupils had been entered for GCSE mathematics in a modular fashion, the school was considered to have a largely linear entry policy.

The sampling frame only included schools that entered 50 or more candidates for GCSE mathematics. It also included only schools whose National Centre Number could be matched to a DfE identification number, for practical reasons. The target population was stratified by school type, proportional to entry; school types were independent, academy, FE college and other school (i.e. local authority). The sample comprised three schools, four academies, one independent school and two FE colleges in the modular group and four schools, four academies, one independent school and one FE college in the linear group. Two replacement samples were constructed.

This yielded 12 participating institutions for the first wave of interviews conducted between April and November 2015. This included; three FE colleges, five academies, two schools and two independent schools. To boost the sample, a second wave of purposive sampling was undertaken, targeting school types that were not represented in the original sampling frame. This led to the additional recruitment of a grammar school.

Overall the sample included 15 institutions (Table 7.1). In the linear sample this included one FE college, three academies, one school and one independent institution. In the modular sample this included two FE colleges, three academies, two schools and one independent institution. A grammar school was also represented.

Table 7.1 Sample by institution type and examination route

Type	Predominant entry policy	No. schools
FE college	Linear	1
	Modular	2
Academy	Linear	3
	Modular	3
School	Linear	1
	Modular	2
Independent	Linear	1
	Modular	1
Grammar	Linear	1
Total schools		**15**
Total interviewees		**49**

Participants

In each institution included in the sample, the head teacher, the head of mathematics and head of English were invited to participate in the interviews. When an alternative had to be sought, the interviews were conducted with deputies or heads of Key Stage 4 in the subject departments. In total 43 interviews were conducted. Occasionally (six interviews) two participants were interviewed together; thus 49 interviewees participated in total. Overall seven head teachers, one vice principal, seven deputy head teachers, one curriculum manager of English and mathematics, 15 heads of mathematics, one deputy head in mathematics, 11 heads of English, one incoming head of English, one teacher in charge of literacy, one deputy head in English, one head of Key Stage 4 literacy and one assistant subject leader in English were interviewed. Time spent as head teacher, or head of subject ranged from a participant who had been in post only since September 2015 to one who had been in post for 26 years.

Interview design and data collection

A semi-structured interview schedule was designed (Appendix K), drawing on the themes raised in the literature review. A list of generic claims in the literature was also generated, to prompt discussion where necessary. Interviewers used the schedule as a guide to discussion, but were free to change the order of questions or to follow up statements made by participants.

The interviews were audio recorded and at the start of each one a general introduction to the project was provided by the interviewer. Here, participants were reminded that the focus of the research was on the approaches to linear and modular assessment in schools, and were asked to answer the questions with a view to three distinct time periods:

a) **before** linear examination changes came into effect (from summer 2014);
b) the **current** system, with all examinations being taken at the end of the qualification, no matter the original route; and
c) in relation to the **future** of the new linear qualifications from September 2015.

Data analysis

Interviews were audio recorded and transcribed. Two researchers independently employed deductive and inductive coding to code four interviews based on the literature and pilot data. They then agreed an outline of codes. This generated six main codes (practices, beliefs about practices, general beliefs about modular and linear, impact of changes, coursework and beliefs about policy changes) and 25 sub-codes (Table 7.2). All interviews were then coded on this basis; coding was conducted by one researcher at Oxford University and two at Ofqual. In the data presented below, extracts are attributed to a job title and a school number, so Head of Mathematics 4 and Head of English 4 are from the same school. Although the appropriate job title for the interviewee has been given, the numbering has been maintained, so for example there is a Vice Principal 4 but no Head Teacher 4.

Table 7.2 Codes and their frequency

Main Code	Sub-code	Sources	Occurrences
Practices	Entry (who & when)	43	142
	Re-sit	40	61
	Examination board	20	52
	Curriculum	26	62
	Managing student perceptions	5	7
Beliefs about practices	Best for students	22	50
	Pragmatic reasons (e.g. league tables)	17	35
General beliefs about modular ad linear	General beliefs	41	240
	Different types of students (e.g. gender, social class, ethnicity, EAL, different abilities)	39	112
	Motivation	22	44
	Linear 'gold standard'	5	9
	Modular provides better data (including external validity, increases student beliefs, changes to teacher perception)	19 16	31 20
	School logistics (e.g. organisation of several examination sessions throughout the year)	18	28
	Maturation effect		
Impact of changes	On student motivation	35	57
	On results - school	37	77
	- individual	16	25
	On school policy and curriculum (KS4 & KS3)	25	64
	On school policy and assessment (e.g. additional 'mock exam' sessions)	25	49
	On examination skills	35	59
	On students' learning (revision and retention)	26	23

Main Code	Sub-code	Sources	Occurrences
	On students personally (psychological, emotional)	22	34
	New grading system	19	30
Coursework		17	35
Beliefs about policy changes	No. of changes	12	16
	Politically motivated	22	35
	Lack of trust in teachers	6	13

Ethics

The study was reviewed by, and received ethical clearance through, the University of Oxford Central University Research Ethics Committee. Participants received an information sheet informing them that that they were not obliged to take part, outlined the risks and benefits of participation and explained what would happen to the results of the research.

Analysis

This section provides an analysis of the data under three codes that produced the richest data and were most pertinent to the research questions:

I. Teachers' views of modular and linear routes
II. Practices and beliefs about practices
III. Responses to policy changes

Examination structure did not change in isolation, thus many of the issues raised related to wider policy changes which, although conceptually distinct from examination structure, might well have interacted with those changes.

Teachers' views of modular and linear routes

Linear gives you time to develop

The existence of a 'maturation effect' with positive outcomes for attainment was described by twelve interviewees in nine schools. Maturation effects were linked positively with linear courses:

> So the fluency of your voice, the articulacy of your voice, your command of language, in fact develops over the two years, and as you mature personally as well your insight matures. So there is really very little benefit sitting any part of the course early.
>
> (Head of English 14)

Maturation was not only seen as a factor in English performance, and one interviewee linked its benefits to having time to increase confidence in mathematical abilities (Head of Mathematics 8). Two Heads of Mathematics (1 and 9), however, did not think there was a maturation effect in mathematics. Further, one deputy head teacher rejected the maturation effect as irrelevant because he considered that the earlier modules of GCSE 'are designed for younger, less mature kids', although he acknowledged that this might have a limiting effect on the difficulty of the material overall. This was also a

consideration for teachers in the further education colleges in the sample, who saw a difference between their adult learners and their students aged 16-18: they considered that linear assessment was more suitable for the adults, who had both the discipline and the maturity to cope with an extended period of unexamined study followed by concentrated periods of exams.

Reform means lower results

There was almost universal belief that the changes to the system were likely to mean lower results nationally; only the grammar and independent school interviewees did not foresee at least a short-term drop in their results. The reasons for this belief were complex, and were not only linked to the switch to a linear structure, although the move to take modular exams linearly was linked to a drop in the national pass rates by one Head:

> I am concerned, and the evidence is last year's national results, where the percentage of students getting [C and above] fell, percentage pass rate in maths fell, although ours increased, and I think it was absolutely inevitable that that would happen, to the point where I assume the people making the decisions wanted that to happen, because it was so obvious that it would.
>
> (Head Teacher 1)

Long term and deep learning

Five interviewees referred explicitly to 'retention' as being key for success in the new linear specification, and a further one referred instead to 'stamina' as a similar concept (Deputy Head/Head of English 10). Linear exams are therefore 'testing a very different thing for, have you got the skills to do that at any one moment in your life, rather than perhaps, are you also, in addition to doing it at one point, capable of retaining it over a much longer period?' (Head Teacher 1). Long term learning seemed to be contrasted with 'cramming' for exams, which could be associated with both modular (because of lack of time between modules, there was not time for slow learning (Head of Mathematics 7)) and linear (in which a final grade 'might not represent necessarily what someone knows but just what they've managed to cram in the last few weeks' (Head of Mathematics 10)).

The effects on motivation

A wide range of views were expressed on the relationship between modular and linear qualifications and the motivation of students, and interviewees could often see both sides of the argument. Early success could be a highly motivating factor:

> We found that it [modular exams] raised aspirations for students, because actually, if they did do really well in an English exam or in a maths exam, or in RE, it raised their own aspirations. The other thing is, we felt that it raised the expectations of staff because they suddenly thought, well, actually, this student can get [this] at the end of year ten, in the November of Year 11, well, actually, they are capable of getting a C or a B in the summer of their Year 11. So, we

> felt it was quite useful for both staff perception and for students themselves.
>
> (Deputy Head 2)

On the other hand, an early module failure was seen to have two possible impacts on students: it could provide a 'reality check' to prevent them falling into 'denial' (Deputy Head 9), but on the other hand it could also be a demotivator, particularly when contrasted against a linear system:

> So I think a lot of students who when we moved from modular to linear - actually liked it, cos we were saying you know it doesn't matter what you get now, we're aiming for a C at the end of Year 11, that's two years away, that's a year and a half away, and actually it can build their confidence, and a few good results here and there can build them up. Whereas if they've got a formal exam that already says you've failed … it was very hard to motivate those students again.
>
> (Head of Mathematics 7)

The movement away from modular therefore 'maybe [did] not built it [confidence] up, but it didn't destroy it' (Head of Mathematics 7). The lack of opportunity to 'put it right' in the case of a failure was identified as a demotivating factor (Deputy Head Teacher 15), but it could also be a good thing, as students in Year 11 were more switched on, and did not have the attitude of being able to re-sit the module (Head of Mathematics 7).

In relation to re-sits in general, one interviewee argued that when 'pupils knew that they could have another chance, they were less focused' or that having achieved a C overall, they would not focus for the re-sit 'because they thought, oh, I've got a C, that's all I need, whereas actually, they're capable of B or an A' (Head of Mathematics 2). On the other hand, regular modular exams provided motivation for the able but unmotivated those who would otherwise 'tend to coast' (Deputy Head Teacher 15).

For some interviewees the modular exams were a key way for students to see the relevance of their learning in relation to their final assessment, to 'see the point in doing this' (Head of Mathematics 9) in contrast to the linear route:

> It's massively overwhelming because they can see how much they've got to cover and you're saying to them 'We're doing this now but you're not going to be examined on it for another year and a half' – and you can see them thinking well you know why am I doing it now, and perhaps actually why is this is important now when my exam's not for another two years
>
> (Head of English 8)

It was argued that the regular stepping stones of modular exams could keep students motivated, when 'there was never an exam more than two or three months away' (Head of Mathematics 9). There were specific concerns for weaker

students in relation to the removal of modular from three interviewees. One felt that students with poor 'maths memory' were better able to concentrate on a short-term goal with an outcome, and 'would feel more comfortable' (Head of Mathematics 1). Similarly, one head expressed his concern that weaker students would be 'overwhelmed' by the content being delivered as a whole course rather than in pieces (Head Teacher 1).

Interviewees in three schools suggested that it was the responsibility of teachers to ensure that the changes do not lead to demotivation of students. One commented that it was important to turn students into 'resilient learners' and to do so 'you've got to make your teachers more positive about come on we can do this and it's okay to fail' (Head of Maths 12). In two other schools interviewees suggested that since students would know no different, their motivation would be unaffected by the change to the new specifications.

Student stress

The majority of teachers who expressed an opinion thought that the linear structure was the more stressful route (10 teachers in 8 schools). A variety of reasons were given, both in terms of the 'weight of final exams' and also the advantages that 'being allowed to track their progress' had for reducing anxiety over results (Deputy Head Curriculum 2). However, two teachers thought that modular examinations made for greater stress for students 'because it's over a longer period, and they're just as stressed no matter what type of exam it is' (Head of Mathematics 2), a point which was also acknowledged by the three teachers who saw stress as a factor in both routes: it's 'an option between constant stress or a huge amount of stress at the end' (Head of English 3).

Is linear the gold standard?

Only one teacher used the phrase 'gold standard' in reference to the linear qualifications.

> Yes, but as a teacher with a strong academic background of teaching an EBacc subject, as a historian, do I think linear's better? Yes, I do, actually, and I guess that's a conflict for me, as a teacher and as a senior leader. I do think that it's the gold standard.
>
> (Deputy Head 2)

Four other teachers suggested that linear was perceived to be 'harder' in reference to the idea of 'rigour', and one suggested 'making them credible to other people, to maybe the public and to potential employers' (Head of KS4 English 2). Two of the four, however, suggested that this view was potentially misguided, and was a nostalgic, rather than evidence-based, view.

Different groups of students

Beliefs about the suitability of modular and linear for different groups of students, such as higher and lower attaining students, varied. Gender was rarely seen to be a factor, and was potentially more likely to be one when linked to student stress, but also likely to be acknowledged as a stereotype. It was suggested that girls might be more prone to stress in the linear system, but boys

were less likely to work hard all the way through in linear. Two or three respondents did suggest that modular favoured girls and linear favoured boys: 'I think the change to linear may, I dread to say, because they will throw me in jail, it may favour boys. I think the previous system is in favour of girls' (Deputy Head 9). Others suggested that the 'goal-focused' nature of boys meant regular modules made them 'pull their socks up' (Head of English 14).

A number of teachers, however, mentioned worries about the effect of the change on economically disadvantaged students:

> Nearly 50 per cent of our students are eligible for the Pupil Premium, and those students, they do struggle with resilience, they struggle with retention of information; they struggle with the stress and pressure of doing a large number of exams in one session, in one sitting.
>
> (Deputy Head 2)

The worry was that 'these changes have been designed with the middle class student in mind' (Head of KS4 English 2), and that this might mean they had a disproportionate impact on results in some schools. Others cited the difficulty of the end of formal teaching leading into a revision period during which children from disadvantaged backgrounds were likely to drop off. It was noted by one teacher that 'by their very nature disadvantaged students are very vulnerable to shifts in their circumstances, and so the one-off all or nothing exam by definition is going to act against them, because if something's happened the night before they are far more disadvantaged than a non-disadvantaged student would be' (Head Teacher 5). Another cited the problems with attendance for Pupil Premium students, which was more likely, she felt, to have an impact on a linear route than a modular one (Head of English 15).

Others suggested that the removal of controlled assessment might benefit those from 'less stable homes' (Head of English 4), and the short intense period of examination would mean 'we can probably do some more work with them' (Head Teacher 4). The similarity of views of different teachers within the same school may suggest that there is some effect on beliefs according to what school practices are, or what the composition of the student cohort is.

Ability was cited as a characteristic of groups who better suited modular or linear assessment. For English, the concern was more clearly evident in terms of the lack of tiering and the amount of reading in the new qualifications, rather than modular or linear assessment, but six heads of mathematics mentioned that they thought that the linear examination structure was suited to higher attaining pupils, and in fact benefited them. 'I think the advantages of linear are for the able students only' (Head of Mathematics 1). Several assumed the converse to be true, that modular suited lower sets, but two respondents disagreed. One noted that in his experience with modular examinations with bottom sets 'we got into a sort of vicious cycle of re-sits and revision sessions and students who were weak anyway being told they were weak again and again' (Head of Mathematics 12), whereas another thought that 'the bonus of the linear is that we can do it in a

year. So I think it's almost top and bottom I'd say linear. Middle, I'd say modular probably' (Head of Mathematics 9).

The effect on students was often not linked to membership of defined groups, however, but to specific individual characteristics.

> It's horses for courses, I think modularity, makes it more accessible for a great many pupils. It's less daunting, and I think that pupils once they go through a linear world will be doing fewer subjects; I can almost see that as being an outcome of it.
>
> (Head of Curriculum 14)

> Some learners might be more akin to having the linear approach and accessing the linear approach in terms of assessment, because they're quite independent in their learning. They've got the stamina and endurance in terms of exams and assessment, end of year assessment in that sense. Whereas others may well benefit from the bite-sized step by step kind of guarding modular approach because they find it's better and they come in every now and then and make a little contribution towards that final outcome than just, you know, having that end product, that end exam. Some of them might not have the stamina to actually be able to represent themselves in the best way. (Deputy Head 10)

Practices and beliefs about practices

Structure of the curriculum and timetable

In most schools, students studied both English (language) and English literature. In five schools changes to the curriculum in English under the new linear qualifications were mentioned. In these schools, there was a move from a system where students would study English literature followed by English language. In the past, schools 'were able to split up what they were focusing on at any one period of time' (Head of English 2) but now they must do both throughout the two years. It was suggested that this can be a benefit, where students were previously completing GCSE English literature in Year 10, if they went on to take English Literature at A level 'by the time they start in September of year twelve they won't have studied literature for the past eighteen months in the classroom in a structured way' (Head of English 3).

Some schools had always taught the English syllabuses concurrently, and one interviewee commented that texts lend themselves to a modular approach, even in a linear system (Head of Literacy 1). One interviewee noted that the closed text approach in the new linear specifications would have an impact on the way the curriculum was structured:

> I think there will be issues in the fact that it's combined itself with a closed text approach. So you've got students studying … texts over two years that they're going to be examined on at the end of that two

> years, so you've got to plan almost like a carousel, a rotation, so that you're looking at everything more than once.
>
> (Head of English 8)

Mathematics teacher interviewees suggested that although the content and order of the curriculum had changed, the pedagogy would not: 'so, the way we're teaching hasn't changed, but what we're teaching, and when we're teaching it' (Head of Mathematics 2). The change to linear examinations might not make a difference to the order of teaching in all schools, however, because as one head of mathematics indicated:

> You teach the easier stuff first and build up anyway because maths being a spiral, you wouldn't teach the harder stuff and then go back, so it is like going through the modules in some ways.
>
> (Head of Mathematics 3)

In one school the top set of students were entered for the GCSE mathematics examination at the end of Year 10, and GCSE additional mathematics at the end of Year 11. However, the change to linear examinations meant that the double entry opportunity would not be offered to the current Year 10 pupils because of the concern that taking both subjects at the same time, at the end of Year 11, could detrimentally affect their results (Head Teacher 1). The grammar school in the sample reported that they would continue to enter all students for both mathematics GCSEs; even those only targeting a C grade for GCSE mathematics.

The time available in the core subjects in schools was varied. In English three hours a week was considered to be unusually low. One head of mathematics mentioned the tension between increased content and the number of lessons:

> We're really struggling to make it fit with the number of lessons we have available to us. Obviously, that's a timetabling issue, but at the head of maths day when we were talking about it, the amount of lessons we have was in the middle of, sort of, 50 different schools. So, it's not that we have the least lessons, so I'm sure there's lots of people in the same boat.
>
> (Head of Mathematics 6)

Conversely, there were seen to be some benefits in the move from modular to linear in terms of the time regained from exam preparation. Teaching on the old specifications in one school was reconfigured so that the syllabus would be complete by February half term (Head of Mathematics 7), with the remainder of the year being devoted to revision. In one school this had always happened, with the curriculum in all subjects being structured to have 'about six weeks of time set aside to revise everything from year 10' (Vice Principal 15).

In one school English gained an extra lesson per week in Year 11, which they used from January to May for doing timed essay practice as exam preparation, once a week (Head of English 4). This school had also implemented additional time for mathematics in Key Stage 3 as preparation (Head Teacher 4). Another

school had considered giving more time to English and mathematics in Key Stage 4 but were 'loath to take away an option subject' to do it, which would reduce their 'broad and balanced curriculum' (Deputy Head 7).

Four schools mentioned either having implemented or planning to implement a two-year Key Stage 3 so that students were ready to begin the GCSE course at the beginning of Year 9 for a three-year Key Stage 4, and a further school had GCSE mathematics teaching beginning at Christmas of Year 9.

One respondent saw the change as a positive one, linking it to a sense of increased autonomy and control over curriculum.

> The modular approach, to some extent, takes autonomy or takes some level of autonomy away from centres and providers and colleges and schools, because you have a rigid set of assessments that are required at specific times to be in place. For me from a pedagogical standpoint that takes the fun and freedom from teaching and learning delivery because of the rigidity of the whole modular approach. With linear I think it gives teachers and education providers the opportunity to be flexible with their delivery to look at covering the important aspect of the specification in place, covering all aspects of specification in place, but having the freedom to provide wider learning as well at the same time without having to do, what I used to call, the stop-start approach.
>
> (Deputy Head 10)

Assessment in schools

Seven of the schools introduced additional mock examinations or other assessments which were intended to provide a variety of benefits previously served by modular examinations or early entry. The reasons given for this included the need to track progress very carefully, in the absence of 'hard data' (Head of Key Stage 4 English 2) which modular examinations provided; giving frequent feedback to the students on their learning; and providing the perceived motivational effect of external assessment upon students.

> Assessment information has a variety of purposes. It can show you how a student's progressing. It also has a value for the student themselves and it acts as a reality check. And that's the bit that's really been lost, and I think that's quite damaging.
>
> (Head Teacher 5)

The mocks were also seen as a way of preparing students to take a heavy examination load at the end of the two years, and familiarity with the examinations which are now so crucial. Introduction of more school mock examinations extended to Key Stage 3 in places:

> So, we are now, from Year 7, giving them the experience of the exam hall, of the invigilators, of the lining up, of the exam timetable, but only in English and maths, and Key Stage 3. So, Key Stage 3 is English

> and maths, and then in Year 10, they have full end of year timetable, and then Year 11, they have full mocked exam timetable.
>
> (Deputy Head 6)

English and mathematics were singled out by other schools for additional assessment; one school which already had mocks in December in Year 11 reported instituting an additional mock in just these two subjects in March. Although for some interviewees mocks were perceived as a 'different type of stress' from the external exams (Head of English 3), a surprising number of interviewees thought that their students would be undergoing far more assessment under the new system:

> It is going to feel to the students a lot more like it is dominated by assessments because we are going to have to, even if it is not a mock period or anything else, we are going to have to assess them in classes quite a lot, to ensure that they are developing those skills.
>
> (Head of Key Stage 4 English 2)

Some stated that one reason to have more assessments was because the mocks were not taken as seriously, so more points for gathering data on the progress of pupils was required (Head of Mathematics 1). The financial constraints of this system in terms of invigilation led one school to choose timed assessments in lessons as opposed to formal mocks (Head Teacher 4) which provides data and feedback, but not the practice at sitting for a long period of time.

Managing external assessments

The logistics of multiple modular examinations was reported as a burden on schools. As well as the paper administration for examinations, there were

- practical issues including finding enough space for students from Year 10 and Year 11 to sit exams simultaneously, and the consequent knock-on effects for teaching space for subjects such as PE;
- the loss of teaching time for module examinations;
- the expense of providing the one-to-one support or other special access arrangements throughout the year instead of for a short time;
- the disruption to teaching in other subjects when students miss lessons for module examinations; and
- the difficulty of moving students between teaching groups on the basis of whether or not they needed to re-sit certain parts of the qualification.

The cost of multiple examinations was also mentioned by three teachers.

Content and difficulty

For English departments the main concern relating to content was the increase in literature teaching. Where previously some schools might have entered some (or all) students for the English GCSE combining language and literature, all students would be taught separate English language and English literature GCSEs from September 2015, because of the double weighting of these qualifications in school progress measures. This became a concern in relation to less able

students 'who would not normally have sat this very hard-core literature style exam' (Head of English 2). The response to this in this case was 'rather than approach the entire texts, we have got a booklet of extracts, so breaking it down, making it more accessible, trying to get to the important bits' (Head of English 2). Others noted the quantity of literature was going to be a source of difficulty: 'there's so much poetry... that 19th century novel as well is massive' (Head of English 4). This might lead to a change in teaching practice:

> There's no way that we can go through an entire Dickens novel page by page – it would take about 10 years ... So … you're teaching a different set of skills, you're teaching them to try and apply the ideas from this extract, because … they've got to be able to be more independent.
>
> (Head of English 8)

For mathematics departments, the amount of curriculum content in the new specifications was more of a concern than the change in structure. Six of the mathematics interviewees commented on the increase in content.

> The only difference is the size of the content. There is far more in the new one, particularly for higher. And the foundation going up to equivalent B means that we have got to get more in for those foundation students than we used to because they don't want to come across questions they have never seen before. So we have to make sure we teach as much as we can to as many as we can.
>
> (Head of Mathematics 3)

Increased content led to challenges in curriculum coverage, but also concern over the speed of teaching needed (Head of Mathematics 6). Some schools restructured the curriculum towards the GCSE content: 'we'll have to use, definitely, all the five years to get them entered at the end of Year 11' (Head of Mathematics 2); or for others maintaining a three-year Key Stage 4 (Head of Mathematics 5). Three schools anticipated entering more students for the foundation tier in mathematics in future because of the increase in content. Demand in the foundation tier curriculum was perceived to have increased, with 'higher grade topics that are kind of B or A grade … are suddenly in the foundation' (Head of Mathematics 7). Schools addressed this change by adapting their Key Stage 3 curriculum. One of the colleges was concerned over the demotivating effect of some the content on foundation tier for those retaking GCSE mathematics:

> If they come across things like trigonometry, simultaneous equations, the stuff they couldn't do in school they're going to be faced with having to try and do that again when they struggled with it previously. So I think they'll get to those type of sessions and just switch off and say, 'I just don't understand it'. It's a far too high level for them to understand, so I think it could be quite demotivating for the maths students with the new content coming in.
>
> (Head of English and Mathematics 13)

On the other hand, some considered that the increased content in GCSE higher tier mathematics 'will probably be a better start to A level for those students who are going to do it' (Head of Mathematics 9).

Changes in examination demands

Removal of controlled assessment in English was one of the biggest changes noted by interviewees; mentioned in all schools. This was seen to have an impact on the skills which were required:

> So skills to do with re-drafting, re-editing and things like that become less important. And skills to do with revision and also to do with building the students' confidence, because an exam is a very independent thing, it is very devoid of teacher intervention. So especially with bottom sets, giving them the confidence to go into a two-hour exam on weighty texts like Frankenstein, and feel they can do it and do all those things, I think that is quite important. So definitely the teaching has had to change in that way.
>
> (Head of English 2b)

In addition, the loss of speaking and listening assessment and controlled assessment was mentioned as removing the requirement for a range of skills from the English curriculum, as these would not be required by a final linear examination.

The nature of the questions in the new mathematics qualifications was also mentioned by five of the heads of mathematics and one head teacher, himself a mathematics teacher. Questions were notably less structured, which no longer led students through the steps they had to take. There was also perceived to be a greater emphasis on problem solving. This had led to a small but significant amount of change in the way they needed to teach:

> But exposing the students to those more problem solving questions, those questions that are linked to other areas, you know trying to get sort of students stuck on a problem and getting them to be resilient to work their way through it, if they don't recognise instantly … you know there's a big danger specially with middle or low ability students of seeing a very wide question and instantly panicking and thinking oh I can't possibly answer that. And so we're entering into our schemes of work far more of those types of questions, and kind of getting students stuck.
>
> (Head of Mathematics 7)

The linear nature of the examination was also seen as having an impact on examination demands. Not only was there more and more difficult content to contend with, students had to 'chunk it down and condense it so that they've got effective revision notes' (Deputy Head 2). Linear exams were a concern to two mathematics interviewees: 'it's harder to guide the students' revision because in linear, they could really assess anything, whereas modular, you can say, 'Well,

there's 20 skills, it's going to be this' (Head of Mathematics 6); and also because in order to address the linear exams 'you need to know everything you've ever learned' (Head of Mathematics 9).

Revision is therefore likely to be emphasised to a greater extent in both subjects, according to teachers. In English the 'perception out there that you don't have to revise for English' was a challenge, because 'they will have to learn quotations, they will have to revise' (Head of Literacy 12). Development of revision skills was also seen as a useful skill for the future; for A level and higher education (Head of Mathematics 2).

Responses to policy changes

This section considers teachers' beliefs about and responses to the changes in examination structure policy, grading, and other changes in the GCSE.

Changes to grading

There was a high level of anxiety related to the move from letter to number grading, particularly regarding the grade C equivalency. One of the interviewees said that the 'new C' but would actually 'equate to an old C+, so that's moving everything up, so that will affect results' (Head of Literacy 1). Additionally the lack of certainty of the relationship between bands in specimen assessment materials and the eventual number grades was a source of worry for teachers.

> What we would like … to see what a level 8 piece of work looks like. I would like a copy, something hard, that says 'This is what a level 5 piece of work looks like'.
>
> (Head of English 8)

In turn this creates issues for students, because they want to know what a mark means in 'real terms', and teachers must answer 'I'm not sure yet' (Head of English 8). One interviewee noted that his students joke that:

> Oh I'll just tell everyone whatever I get, 'that's equivalent to an A*.' [laughs] Which you know you can see actually for some employers, they would be like 'Oh what's this 5?' 'Oh it's an A*' 'Oh okay that's fine' – they might not understand it.
>
> (Head of Mathematics 7)

The effect which the changes in grading would have at the top end was also a source of speculation for interviewees, in particular the split of the A* into two grades, which might leave schools with a lower number of 'high A*s' with a small number of top grades (Head of Mathematics 2). Another interviewee suggested that 9s might be very rare indeed (Head of Mathematics 12).

Perceived lack of trust in teachers by policy makers

Five interviewees identified a sense that the policy changes had been made in response to a lack of trust in teachers on the part of examination boards and the government. The link was specifically made to the abolition of the speaking and listening component of the English GCSE and the removal of controlled

assessment; it was noticeably not raised by any of the mathematics teachers. One interviewee identified the challenge to her identity as a professional that assessment was 'being removed from us' because of 'a sense of distrust that we are just fiddling the grades and cheating' (Head of English 2). Another commented that this was 'not great for morale within the profession' (Head of English 6).

Policy is to lower national results

Six interviewees suggested that the changes in policy were designed to lower the pass rate at GCSE to combat grade inflation, and a perception on the part of policy makers 'that the exams have got easier over the years' (Deputy Head 7) or that the modular structure was 'making it too easy' (Head of Mathematics 5). 'We know it's politics, I suspect it's because the league tables are not showing the results that the politicians think they should be showing because some schools are playing the system' (Deputy Head 9). There was also a suggestion that the new grading system would obscure the change and make comparability year to year over the beginning of the specification hard to see, which was described as a 'cynical' move by one interviewee (Head of Mathematics 3).

Policy is to make us more competitive

Linked to the belief in the desired drop in national pass rates was the related concept of 'rigour' and increased academic rigour being linked to linear qualifications. This was seen in two ways in the data: as a way of proving to employers that the education system was fit for purpose, that exams were 'credible' (Head of Key Stage 4 English 2), and in terms of international comparisons, both within Europe and more widely.

> What do you think is behind the policy? It is to be more competitive... in the market and there are problems with maths and science that Britain is not doing too well and lots of the Asian economies are getting much better. So I can see some of the rationale behind it, yes.
>
> (Head of Mathematics 10)

The level at which the good pass had been set in the new system was suggested as being 'to bring us in line with the top performing countries in Europe', although 'time will tell' (Head of English 3). Another interviewee challenged the idea that we needed to be keeping up with other countries that might be 'testing very different things' (Head of English 6).

Perceptions of the policy process

Twelve teachers considered that the policy changes had been guided by the personal views and educational experiences of the former Secretary of State for Education, Michael Gove. In particular, eight of them suggested that the new system was largely based on 'his recollections of what his school days were like' (Head 5) with 'old style exams' (Head of English 3). One teacher suggested that the policy was based on an underlying belief about assessment, that 'an exam and a qualification should be what you know at that point about everything' (Head of English 2). Seven interviewees suggested that there was a need for greater understanding on the part of policy makers of the need for education

policy to encompass the needs of the 'average person' as well as those who are 'able to go through at higher levels' (Head of Mathematics 3).

The need for a period without change

Thirteen teachers in nine schools felt that there had been a period in which huge numbers of significant changes had occurred, and made a plea for a period of 'time to just embed these things' saying, '... it's very hard to keep up with' (Head of Key Stage 4 English 2). In particular, the fact that reform had come at the same time for Key Stage 3 and Key Stage 4 (and also Key Stage 5 in English) made it difficult for departments who felt 'all a bit rushed and pressured' (Head of English 6). Two teachers identified 'pendulum swinging' in education policy (Head 1) which was related to 'change of government or change of education minister' (Head of Mathematics 8). Teachers are keen for a 'period of stability, where there are no changes' (Head of Mathematics 1), and for that period to be counted in years. The commitment to no more in-year announcements of policy changes was 'still not a lot' (Head 1).

Summary of baseline research in schools findings

What do teachers believe the effects of examination route are upon overall GCSE outcomes and grade inflation?

As the reforms involved changes to the examination structure in combination with increased curricular demand and more challenging examinations, findings regarding examination structure were confounded in the research. Teachers believed that the changes were designed to combat perceived grade inflation and to reduce rising GCSE outcomes. In mathematics, there were concerns that the curriculum demands and increased curriculum could cause difficulties for curriculum coverage in timetabled slots.

What do teachers believe the effects of examination route are for different groups of pupils?

Linear examinations were considered to be better for more mature (generally older) pupils. Weaker students were thought by some teachers to be better suited to modular syllabuses, especially if they had poor memory for the subject. In mathematics in particular, linear assessment was thought to suit stronger students more. There was some agreement with the acknowledged stereotype of girls suiting modular and boys linear. Modular examinations were deemed to have motivated students who had done well on early modules, but for the group of students who did poorly on these, it was observed to have been demotivating. For some groups of students, such as disadvantaged students, it was reported that linear examinations might be better because of the concentration of assessment over a short length of time, so that they were less likely to forfeit sections of the qualification; however, absences during the year might well have an impact on final success. Views were varied about the impact on Pupil Premium students, but concerns were raised that linear qualifications, and the changes to the system in general, would disadvantage these students further.

How, if at all, has the examination route influenced teaching strategies and teaching materials according to the teachers?

Somewhat counterintuitively, introduction of linear examinations meant a consequential increase in school assessments, with teachers conducting more mock examinations. There was a shift to teaching revision and examination skills to a greater extent, as the stakes in the linear assessments were higher. Teachers recognised the benefits of the shift to linear examinations because they would have more teaching time with fewer examinations scheduled. In line with policy aspirations, teachers indicated that they would be teaching different skills, such as more problem solving. Further, it was thought to be less easy to predict the examination questions and focus teaching upon those areas in a linear examination structure.

Some large changes to the way in which curricula in schools were structured were also found. For example, in some schools English language and English literature would now be taught concurrently rather than sequentially. Additionally, some schools had changed the duration of Key Stage 3 to two years and Key Stage 4 to three years.

Chapter 8 Research in schools: linear examinations phase

This chapter encapsulates the key findings from the second wave of qualitative research conducted with key participants from institutions across England about how the embedding of linear English and mathematics GCSEs has impacted on the way they plan and teach the GCSEs. The rationale behind this second phase of interviews was that we would be able to not only collect teachers' perceptions of the linear and modular approaches, but also record what schools had done or were doing in response to the policy change. As a reminder, this phase of the project investigated the following research questions:

1. What do teachers believe the effects of examination route are upon overall GCSE outcomes and grade inflation?
2. What do teachers believe the effects of examination route are for different groups of pupils?
3. How, if at all, has the examination route influenced teaching strategies and teaching materials according to the teachers?

Method

Sampling

In the first phase of qualitative research, the participating schools were selected using a random stratified sample divided equally between institutions in England who entered pupils for modular and linear qualifications in mathematics in 2013. If the schools entered 50 per cent or more of their GCSE mathematics pupils for a modular qualification in 2013, they were classed as having a modular entry policy. Those that did not meet this requirement were classed as having a linear entry policy. Care was also taken to ensure that the sampling only included those schools that entered 50 or more candidates for GCSE mathematics, except in the case of special schools which we were unable to recruit in this phase. This led to a sample of 15 institutions where interviews were conducted between April and November 2015.

In this second phase of the project, the same set of 15 institutions was approached for interviews to understand how the change from modular to linear examinations had affected the institutions (Table 8.1). Where institutions were unwilling or unable to take part, appropriate replacement schools were approached for recruitment. Effort was also made to broaden the school types to include a special school, where we were successful. The institutions were approached for interviews beginning in May 2017 and interviews were completed by November 2017. This recruitment round yielded a sample of 12 institutions. They are shown below, along with their predominant entry policy in 2013.

Table 8.1 Sample of institutions by institution type and examination route.

Type of institution	Predominant entry policy in 2013	No. schools	No. interviewees
FE College	Linear	1	3
	Modular	2	5
Academy	Linear	2	6
	Modular	2	6
School	Linear	0	0
	Modular	1	3
Independent	Linear	1	3
	Modular	1	3
Grammar	Linear	1	3
Special Schools	Modular	1	3
Totals		**12**	**35**
No. of interviewees common to both rounds of interviews			21

It is important to remember that, although the purpose of the second wave of research was to capture perceptions of the change from modular to linear, the examinations sat in the summer of 2017 also involved new, more demanding content[38] and new grading (9 to 1) structure. Final re-sit opportunities were also being offered for legacy GCSEs at this time, a scenario relevant for at least two institutions in our sample set. It is also key to note that International GCSEs in English and mathematics were also still available for examination that summer, in both grading structures: A* to G (non-accredited, and not counted towards performance tables); and 9 to1 (regulated by Ofqual, but not counted towards performance tables). Where such International GCSEs were still being offered at our interviewed institutions, questions were posed to determine future changes that would be made when moving to the new linear GCSE specifications. The key changes to the GCSE syllabuses for both subjects have been captured below:

English

- removal of GCSE English qualification which combined language and literature, now separate GCSE English language and GCSE English literature are available
- removal of all controlled assessment
- removal of tiering (used to be foundation and higher tiers)
- spoken English assessed in a formal presentation against common criteria with a separate grade (pass, merit or distinction) reported as an endorsement to the qualification grade
- English language unseen texts cover 19th, 20th and 21st centuries, and a range of genres and types but transient texts e.g. instant news feeds not included
- English literature set texts no longer include: American novels (e.g. Of Mice and Men) as fiction, drama from the British Isles from 1914, multicultural poetry as all works must have been originally written in English (as a standalone; there are one or two remaining)

[38] https://www.gov.uk/government/publications/get-the-facts-gcse-and-A level-reform/get-the-facts-gcse-reform

- set texts now include a compulsory 19th century British novel (most students previously studied a 20th century novel), a 20th or 21st century (1914 onwards) British play or novel, romantic poetry
- Shakespeare remains compulsory
- exams are now 'closed book' (where mostly clean texts were allowed previously) but extracts are provided in the questions/resource materials for the assessment of textual analysis

Maths

- some content which was previously only on the higher tier is now on foundation tier, along with some additional more demanding content in the higher tier that used to be assessed at AS level
- some formulae which were previously given to students in the examination booklet now require memorisation
- a greater emphasis on problem solving
- a greater emphasis on mathematical reasoning

Participants

At each institution, the head of English, head of mathematics and the head teacher were invited to participate in the interviews. When alternatives had to be sought, the interviews were conducted with suitable replacements, such as the head of KS4 in each subject or the deputy head in charge of curriculum. On occasion, individuals were interviewed who served in dual capacity, for example, head of mathematics and deputy head of curriculum (three interviews). In total, 35 interviews were conducted. This involved conducting interviews with five head teachers, two deputy head teachers, one deputy head in charge of academics, two deputy directors of curriculum, one head of studies, one vice principal of mathematics, 11 heads of mathematics, 13 heads of English, and two directors or managers of English and mathematics.

Data collection

A semi-structured interview schedule (Appendix L) was designed following the implementation of the linear GCSE English and mathematics, drawing on understanding that was generated in the first wave of qualitative research in 2015. Interviews were carried out by six different members of the research team from both Oxford University and Ofqual, and the schedule was provided as a guide to discussion. Interviews at three institutions were held before the GCSE examinations began, 10 were conducted during the examination period, and one was held after the publication of GCSE results.

The interviews were audio recorded and at the start of the interview, a generic introduction to the project was provided by the interviewers. Here, the participants were reminded that the purpose of this research was to see and understand what changes have happened at different institutions due to the change from modular to linear GCSEs.

Data analysis

All the interviews conducted were audio-recorded and transcribed. Following transcription, two researchers from the team coded six interviews using both inductive and deductive coding techniques. The data was coded to the three distinct research questions. After independent coding, the two researchers agreed on an outline of codes to generate a coding framework for the whole dataset. This coding framework was then used to code the whole dataset: coding was carried out by two researchers at Oxford University. Schools in the dataset were assigned a distinct number and extracts from the data reference the individual by job title and assigned number, for example, Deputy Director 7.

Analysis

Given that the move to linear exams has also involved the introduction of a new, more demanding content in the exams, findings regarding the examination structure have been confounded in this second phase of research. Efforts were made by the coders to separate the examination structure and only code for linearisation. There were, however, instances where it was hard for participants themselves to separate changes to the curriculum from changes to the structure of the examinations.

What do teachers now think of the linear examinations?

As part of this research we were keen to understand whether teachers' opinions of the linear examinations had changed over the last two years, while the new qualifications were being embedded, and to try to understand how and why:

> So, it's easy to sit here and just say it doesn't feel right, and it feels like an unnecessary change, but actually, I'm not closed to the idea that things may emerge from that process and system that actually we recognise as a value.
>
> (Head Teacher 9)

Overall, teachers held mixed views on the new linear examinations. Twenty-one interviewees held an overall positive view of the new linear examinations in that the examinations provided a level playing field, allowed teachers to teach content in a more comprehensive manner, and brought back a sense of fairness to all the students. Eight of these teachers said that they had become pro-linear over recent years as they experienced the new courses. This idea of fairness was highlighted by them, as seen in the school below through the notions that while the intention of the modular courses might have been to provide help to those who really needed it, the flexibility of the systems had allowed for some schools to take advantage of it:

> I've always been very much, from the even playing field perspective, I am very much in favour of a linear course anyway, always have been. I think it's too easy to manipulate the system otherwise which, although we don't do, when you hear about everybody else doing it,

> it's just so unfair and you just can't manage it. So, no, I'm all for this and the 100% exam as well.
>
> (Head of English 9)

Four interviewees also expressed the view that they were pleased to see the removal of controlled assessment as the rising difficulty of the controlled conditions had made the assessment feel artificial (Head Teacher 7); the controlled assessment felt more like a 'teacher proof reading exercise' (Head of English 10), rather than actual teaching of content; and teachers also felt that some manipulation of student performance had been allowed through multiple re-sitting of modules (Deputy Head of Academics 5). The removal of controlled assessment had also been welcomed by a few teachers in the first round of interviews as they felt that the constant cycle of controlled assessment had prevented them from teaching the course in the manner they would have liked:

> When talking about modular courses again, it might not be the system, but it was just how students were reacting to it, but we ended up getting this treadmill of revision, exams, re-sits, revision, exams, re-sits, and actually that takes away from teaching; that takes away from students learning, actually, because they're just always prepping for another exam.
>
> (Head of Mathematics 9, Phase 1)

This was echoed in this second round through some teachers' beliefs that the linear examinations were better because they now had the freedom to teach the whole course holistically, and allowed them to try and cultivate a deeper learning in their students and prepare them for real life (Head of English and Mathematics 8). One teacher even spoke of how the linear system encouraged students to work harder which would prepare them to be more competitive on an international level (Director of English and Mathematics 8).

This sense of fairness and freedom stands in contrast to the first round of interviews where five teachers had expressed a sense that the policy changes had been made due to a 'sense of distrust that we are just fiddling the grades and cheating' (Head of English 4, Phase 1). Where teachers in the first round had been more vocal about their opinions of the policy changes, during the second phase the focus was more on adjusting to the changes and moving forward. Most of the teachers in this phase recognised that the move to linear examinations needed to be managed correctly in terms of its content to make it a successful move, and to ensure accessibility to all students as highlighted by one teacher:

> I don't mind that at all. It's just got to be done the right way, with the right content. This is going back 30 years, isn't it? We're in a cycle, and it's not too problematic – it just needs to be managed really well in terms of what we're asking students to do. Whether we ask them to be examined at the end of the course as opposed to other times, essentially, yes, I don't think matters to them or to us – what matters is what we're examining and how we're examining it
>
> (Head of English 5)

Nine interviewees felt that this transition and the content introduced had not, in fact, been managed correctly, thereby making it harder for students to focus on the intended deeper learning, as their focus was now more towards honing their memorisation skills. Some teachers felt that the content was too difficult and too large for it to be assessed by students in a fair manner through examination at the end of two years.

> I think it's harder for very weak students to learn quotes but do you know they were hammered into them and I don't know of any bottom set that couldn't give you at least five quotes from their Shakespeare play. You know, as simple as they probably were, they probably used those five quotes in any essay but that's naturally going to be harder for them because to memorise is a, you know, is a skill that they probably haven't got.
>
> (Head of English 11)

Also, teachers, while welcoming of the idea of 'levelling the playing field', were of the view that the linear examinations were not fair to all students, particularly those whose prior attainment was not the strongest (Head of English 2), as well those of who had learning difficulties (Head Teacher 7). This concern had also been expressed by the teachers interviewed in the first phase of this project. Teachers in this second phase of interviews alluded to the higher levels of stress that the students were feeling, as well as the impact the examinations were having on their self-confidence and motivation. One teacher, for example, argued that where the new examinations could really give students a sense of what it means to be a true mathematician (in the application of the tools holistically), the difficulty of the examination had him 'worried that it's going to cut the number of students in further maths' and that his 'colleagues in other schools worry about the same thing.' (Head of Studies 6)

This sense of fairness was also emphasised when talking about students with special educational needs and disabilities:

> You know we have to have exams that cater for all students, not the highest ability. You know we can't have a system that causes a whole bunch of kids to fall by the wayside. We have to support them and provide them with a qualification that's meaningful and it feels really quite cruel when you're forcing students with quite severe learning needs through a qualification that actually they can't access.
>
> (Head of English 2)

Despite the phased reform of GCSEs and A levels, six interviewees were critical that the changes to the GCSE examinations had all come too quickly and that there should have been a staggered approach, allowing for teachers and students alike to cope with all the changes and adjust to them.

> No, I think that's the main thing, give us time to get these kids used to what the changes are going to be before we chuck the changes at them, you know, it just always seems so unfair to me. And there's some year groups that become guinea pigs year on year. I mean obviously they're the first cohort through on the new GCSE, they'll be the first cohort through on the new A level and we're still trying to get to grips with it ourselves and then we've got to teach it, give us more time.
>
> (Head of Mathematics 11)

What do teachers believe the effects of examination route are upon overall GCSE outcomes and grade inflation?

As stated before, the reform to the structure of the GCSE examinations in mathematics and English has coincided with curricular change and a more demanding examination, as well as a change to the grading system i.e. a move from A* to G grading to 9 to 1 grading. This is particularly apparent in the entangled findings where teachers were specifically asked about how they felt student performance would be affected in the new GCSEs.

Performance

The majority of teachers interviewed were of the belief that the performance of the school would drop in the summer's GCSE results. This was not the case, however, with those teachers who were interviewed at the grammar or independent schools. The reasons given for this perceived drop in performance were varied and complex, and not solely linked to the switch to linear examinations.

> I think whether it's modular or linear the courses are harder now and it'll be interesting to see what, you know, the people you're interviewing are saying, but if I look at the history, the history I'm teaching now, it's not just O level, it's O level plus. There's no doubt, it's unnecessarily hard and complex and out to catch people out I feel, so yeah I expect our results to go down.
>
> (Head Teacher 7)

Ten of these teachers were resolute in their belief that the change to a purely examination-based measure would definitely have an impact on performance, and not just in the short term, but also in the following year where all subjects would be switching to terminal exams.

> We will have to see – I mean, it's interesting at the moment, it's just English and maths going through this year – now, a lot of parents and students place a great deal of emphasis on English and maths, and will sometimes sort of sacrifice their learning in other areas to prioritise what they're doing in English and maths. Now, we will see changes, I'm absolutely sure, with the results, when the English and maths results come through, but I wonder, in a year's time, when geography and history and science, and everybody else are going

> through the new courses, I wonder how the students will cope with subjects where all the coursework disappears, where there's all that exam learning that's going on, and you can't help but feel that there are certain types of students and learners that'll be really disadvantaged by that.
>
> (Head Teacher 9)

As discussed in more detail below, some teachers were also concerned that the performance of the lowest and middle 'achievers' were going to be affected the most negatively, as they 'will find it harder, and would have benefited from the coursework and things on the old exam' (Head of English 12).

> Anything that's purely exam based, which is what we're talking about, that sort of exam based assessment for most subjects, is going to disadvantage some students, and lends itself towards a certain type of learning, a certain type of learner, and I think it will have an effect, it will have an impact.
>
> (Head Teacher 9)

Some teachers on the other hand spoke of the 'the flipside of [the] all or nothing' nature of the terminal exams, was that it gave some students time to mature and be ready for examinations, rather than having been told 'two or three times that they were failures' (Head of Mathematics 9). Twelve teachers at seven different schools were of the belief that the results should not change dramatically if the 'teaching underneath is still good', as that would mean that students would be prepared for any movement of the goalposts (Vice Principal in Charge of Maths 4).

Grade equivalency

One cannot discuss the above-mentioned perceived drop in performance without talking about grade equivalency. As highlighted previously, the new GCSE examinations were also accompanied by the introduction of the new 9 to 1 grading system. In the lead-up to the examinations, some teachers struggled with a lack of clarity on what the new grading structure meant and how that compared to the old A* to G system. This made it very difficult for schools to predict their students' performance. The majority of teachers held the belief that, while schools that had been exclusively modular and had made the switch to linear examinations might feel a drop in performance, nationally, 'the picture should come out much the same' (Deputy Head of Academics 11). Attention was also drawn to this by 10 teachers through a conversation about the boundaries for grades, and how they could be adjusted to maintain outcomes for the new examinations:

> Boundaries can move, etc. so that what you see from year to year is this idea that actually we've got the same sorts of students every year. So I think to some extent what will change? Well I don't think a lot will necessarily change because I don't think it's going to be in anybody's best interests to suddenly turn around and say 'We've put in these new exams and everybody's failed at them'.

(Vice Principal in Charge of Mathematics 4)

Concerns were expressed that not having any equivalency between the results this year and previous years would result in not only demotivation of students, but of teachers as well (Head of English 7). One teacher talked about how the specifications for the examination would have to move back closer to the older specification because otherwise examination boards would find themselves in a situation where the examinations would seem almost unsuitable for the population that had taken them:

> Um, so I just predict that while they'll still have that problem solving emphasis and the new concept will still be there, I just think the marks that we have seen and some of the national data I've seen from Edexcel, from mock papers, that they've got the data back from schools are so low, almost embarrassingly low that something's got to change otherwise the exam doesn't feel like it's got any validity.
>
> (Head of Mathematics 2)

It was also pointed out by some teachers that whilst the move to linearity would be harder for some students, and the new examinations were harder, the national picture would not be impacted as much, due to political reasons.

> Well, I hope so, but I doubt it, because to have an effect on results, the pass rates would have to go down. If they are everything they claim about being harder, etc., then pass rates will go down. No government's going to allow that, so they won't change anything.
>
> (Deputy Head 5)

Fairness

Ideas of fairness were highly salient with teachers talking about the linearisation of the GCSE examinations. They approached this notion in varied ways however. Seven teachers shared the view that linear exams brought back a sense of fairness to the system, allowing for all students to be tested at the same time with the same set of exams. This was made particularly obvious when these teachers highlighted the challenges of the modular qualifications, such as the different levels of control around controlled assessments at different schools, or the potential for abuse with the availability of retakes. There was also a sense that the modular examinations were becoming almost too predictable from the students' and teachers' perspectives:

> I mean one thing I would say it's probably going to bring a fairness back to the system, because you don't … you know if people are doing coursework in other schools and you start … you just don't know.
>
> (Head Teacher 2)

Some teachers also felt that the system of re-sits and modular examinations, while good for those students that needed more help, was unfair on those students who had higher prior attainment. Teachers also acknowledged that

while re-sits seemed like students were allowed to 'cherry pick and do what [they wanted] to do' (Head of Mathematics 2), and that in some cases, modularity and the ability to re-sit examinations did not really help students learn content effectively (Head of Studies 6).

> As I said, I certainly had issues with the idea that you could keep re-sitting modules. That, to me, is – it's great for them that need to, but it's unjust for them that don't. But as usual it's that phrase, you know, ah, but they're all right. Well, 'they're all right', to me isn't good enough. Every child should be somehow stretched and challenged. You shouldn't ignore the clever ones just because they're clever.
>
> (Deputy Head 5)

One teacher considered that whilst the linear examinations seemed fairer overall from a system point of view, on an individual level, they felt very unfair to the children themselves, as the examinations felt like they were more about memory skills than content:

> So I guess I thought linear would be fairer. With hindsight, I don't think it is. I think there's an awful lot to remember. I think we're examining on memory too much, which isn't necessarily fair because it's not about what they remember, it's about what they can apply and their skills. I worry that we won't be testing that as much now because they're going to be too concerned about what they've got to remember.
>
> (Head of English 12)

This sense of unfairness was also echoed by other teachers in that the linear examinations took away the opportunity to tailor their approach for the diverse groups of their students (Head of Mathematics 2). Having modular examinations had not only allowed teachers to help set realistic expectations with students and parents about performance (Head of Studies 6), but also allowed for some students to sit an examination earlier more strategically and assess if more work was needed to better their performance, and therefore take the 'pressure off them and then if they don't pass they can have another go, a bit like a driving test' (Vice Principal Head of Curriculum 4). This was particularly highlighted through the needs of those students who needed more help or time:

> Yeah. But I think overall, I think of the two systems, a system that had different ways of assessing seems fairer than a system that is completely linear and completely based on the same method of assessment because if your strength is not that, for example, if you're dyslexic then you can't succeed.
>
> (Head of English 4)

Raising standards

Changes to the GCSE examinations were motivated by the Government's commitment[39] to raising standards in schools. The teachers interviewed were divided in their opinions on whether these changes had indeed helped raise standards. Eight teachers expressed their concern about whether the government had deployed the right strategy in trying to accomplish this:

> I don't think much consideration has been given by anybody in government as to what that actually looks like for a 16-year-old
>
> (Head of English 4)

One teacher suggested that 'for the first five years there will be a decline in standard[s]', as schools would have to offer fewer GCSEs to provide teachers and students with enough time to prepare for the changed examinations (Head Teacher 12).

Some teachers believed that standards would not be raised as not enough time had been given for schools to prepare and implement all the changes effectively – 'I don't think let's raise the bar tomorrow is great strategy to go for because I think raising the bar is something that you, … need to take time to realise' (Vice Principal in Charge of Mathematics 4). One interviewee suggested that with all the implemented changes, it felt that rather than moving forwards the examinations had taken a step backwards, to something that had already been tried and tested and failed, particularly in mathematics (Deputy Head 5). Another teacher even commented that, whilst there may have been a desire to 'produce independent thinkers', the nature of the linear examinations was actually resulting in a lot of 'spoon feeding' of the students through exam skills preparations (Head of English 2).

Not all teachers interviewed were as critical of the Government's drive to raise standards. Some viewed the changes quite favourably. These teachers felt that the reforms had put schools on the path to these higher standards. Four teachers explicitly singled out having the opportunity to 'encourage a lot of deep learning in every lesson' (Head of English and Mathematics 8). One school tried raising the bar by encouraging more students to take the higher tier in mathematics (Head of Mathematics 9) to really challenge their students and explore their potential. In English, two teachers at different schools spoke of 'Year 9's [upping] their game' (Head of English 9), as they had made Year 9 more rigorous to prepare them for the incoming GCSEs.

Restrictive curriculum

Five interviewees described the new examinations as restrictive. They were very vocal regarding the fact that they had to spend more time on a more 'traditional curriculum' as there was a push towards the measures of the English Baccalaureate. They felt this meant that there was a more restrictive curriculum being championed, where 'there's much more resources now in the core and in the EBacc subjects and you know things like design and technology very rapidly

[39] https://newgcses.campaign.gov.uk/

are … disappearing' (Head Teacher 2). These teachers believed that the instituted changes had not considered that some students may also need a curriculum that could help them on a path to an apprenticeship or a job, rather than just being targeted towards those students who are 'Oxford, Cambridge material' (Head of Mathematics 12).

> The bit of the literature that presents a problem … Sometimes, it's the language of literature is hard for them. But, I think, if I have to be brutally honest, they can't see how literature can get them a job.
>
> (Head of English 1)

What do teachers believe the effects of examination route are for different groups of pupils?

For the most part, teachers were hesitant to make predictions about the impact of the linear examinations on student performance, however the data presented below represent the different groups of students discussed. This was in part due to the difficulty they had in separating the impact of linearisation from changes in content, and how all the changes were going to have a combined impact on performance. Whilst beliefs about performance and the suitability of the linear exams for distinct groups of students varied, there were some key themes which emerged, which have been highlighted below.

> The interesting thing is that yes, different types of students, we can no longer predict who's going to do well on this exam. Until I can see the results in the summer, I'm not sure who's going to come out on top.
>
> (Head of Mathematics 10)

Gender

In general, gender was not highlighted as a major point of difference in performance. Some anecdotal and personal experience was presented, however, regarding differences in levels of anxiety between the two genders. There was a general acknowledgement among the teachers that girls were more anxious about the terminal examinations than boys (Head of English 2), and would have felt the stress of that final examination more than boys (Head of Maths 7):

> It's girls that seem to somehow over think the process or beforehand really get quite stressed about it, so I think that's, I think terminal exams are probably okay for most of our boys.
>
> (Head of English 2)

Seven of the respondents suggested that the linear examinations would favour boys over girls. 'My personal opinion is that, and perhaps this is gender biased, but boys, in particular, perform well in exams. They like being able to work intensively, under pressure, and often they produce their best work that way' (Head of English 6). Two interviewees also commented that the increased curricular demand and change in expected skills required for the examinations required them to work harder with boys than girls:

> Yeah, so … so a lot more work on concepts of extended writing and we're a very boy dominated school. So we're having to do a significant amount of that. Um, and also you know the levels that they were having were based on reading, not on writing and a lot of boys can be weak writers even though they're quite high ability.
>
> (Head Teacher 2)

Pupil Premium

Several teachers expressed concerns about the performance of those students from economically disadvantaged backgrounds in the new linear examinations:

> But it seems to be that the students that found this the most difficult change are the students who come from families and sort of backgrounds where they, for whatever reason, there are other priorities within the family than education. So their attendance isn't as high, they don't have a place to work at home, there's not that cultural capital, there's not sort of this wider awareness of things that sort of, you kind of need to be able to approach the unseen reading papers with real success.
>
> (Head of English 4)

Some teachers felt that the 'all or nothing culture', created by the 'high-stakes final, terminal exams' (Head Teacher 10) would have a disproportionate impact on Pupil Premium students who wouldn't have the best attendance (Head Teacher 7), supportive family structures (Head of English 9), or the cultural capital (Head of English 12) that would be needed. One teacher commented that 'no matter what qualifications you're going for, resilience is key', and when you're a Pupil Premium student, 'there is less resilience due to less support at home, and having resilience in short supply makes them [examination questions] just that bit more inaccessible' (Vice Principal in Charge of Mathematics 4).

> I think realistically the expectation on students in general to take on the length of qualifications they are now doing seems a little bit crazy, it seems a bit over the top. If again taking your student who is generally not going to do very well in any type of maths exam that they sit, to force them through four and a half hours' worth of exams that they're genuinely not going to do very well in, does nothing for their self-esteem and having gone into one and felt that they've not done very well in it.
>
> (Vice Principal in Charge of Mathematics 4)

Teachers also spoke of interventions that they were putting in place to help these students such as sessions after school or on Saturdays and mapping out daily activities for revision. However, it was reported that there was little to no uptake from this particular group of students among whom self-motivation was not high (Head of English 4). Teachers also spoke of trying to build students' vocabulary in order to be able to access the language paper, as they may not have the 'breadth of reading' behind them (Head of English 9).

There was also a number of instances where teachers spoke of how they had worked hard to close the gap between the Pupil Premium students and the non-Pupil Premium students, and were no longer able to detect a gap in performance between the students.

> We're lucky in the sense that when we collect data three times a year so one of our assistant heads looks at all the data, gives it to us and our pupil premium achieve slightly better in maths than our non-pupil premium. Everything we do is the same for every student. Yes, you know, we've got the small classes for everybody. If a child needs more support we'll then put staff support that way.
>
> (Head of Mathematics 12)

One teacher also expressed concern that while so far they had not seen any differences in performance and suitability among the Pupil Premium students, this may not remain the case in future, when all examinations switch over to the terminal structure (Deputy Head of Academics 11).

Prior attainment

The teachers interviewed believed that all students at their schools would be affected by the changes to examination structure but in different ways, positively and negatively. In the case of those students who were the lowest attaining, several teachers expected to see a drop in their performance partly due to those students being unable to take bite-size exams (Head of English and Mathematics 8). The lack of accessibility associated with the more challenging examinations was also a concern for the lowest attaining students, the impact of which was further compounded by the changes to the curriculum and to the tiering of examinations. Modules, and therefore modular exams, were considered better for lower attaining students (Head of Mathematics 5), and there was a growing concern that these students would find it very challenging to cope with all their subjects when GCSEs were 'pure linear in two years' time' (Deputy Head Academic 11). In the first phase of this research, numerous interviewees stated that modular examinations had been used to motivate and engage students as there was never an 'examination more than two or three months away' (Head of Mathematics 9, Phase 1), and that early success would 'raise aspirations for students' *(*Deputy Head 4, Phase 1). This was reiterated in this phase of interviews where some teachers expressed concern with the rising disengagement and lack of motivation that students were exhibiting due to the 'hardness' of the new examinations (Vice Principal of Curriculum 4).

On the flip side, most teachers felt that the higher attaining students had risen to the challenge, (Head of English 2) and would perform well in the linear examinations but they suffered from anxiety due to the pressure to perform well:

> We have a number of very high-achieving students. A significant proportion of those are girls. Some of those girls are very driven, and actually when they're anticipating getting high grades in English, and in a formal assessment, they maybe get a 3 or a 4, well below

> their expectation, I think they've lost confidence. I think it's caused them great anxiety. It's felt as though there's been a lot of … anxiety and worry.
>
> (Headteacher 9)

Some believed that linear examinations were fairer for the higher attaining students as modules allowed for all students' results to be better (Head of Mathematics 5), rather than allowing the higher attaining students to shine. Some teachers also believed that the modular examinations sent a negative message to students in the lower attaining brackets due to constant re-sits and revision (Head of Mathematics 9). Seventeen teachers, however, believed that the linear examinations were unfair to those who were lower attaining due to the inaccessibility of the examination itself.

> I think we'll find nationally that results will be lower. I think what they've done is, you know … to an extent I can see why they've gone to a linear framework so that if you're looking at the very top performing students, you know, you are really looking at are they Oxford, Cambridge material. But what it doesn't take into account are those at the other end where actually, we'll have quite a few children that won't score.
>
> (Head of Mathematics 4)

SEND

Once again, fairness was a theme when teachers discussed students who might have special educational needs. Teachers felt that the linear examinations were not only unfair to lower attaining students but even more so to students who had special educational needs:

> Yeah. But I think overall, I think of the two systems, a system that had different ways of assessing seems fairer than a system that is completely linear and completely based on the same method of assessment because if your strength is not that, for example, if you're dyslexic then you can't succeed.
>
> (Head of English 4)

This was particularly stressed at the special school, as it was argued that the linear examinations no longer provided their students with the flexibility that some of them needed to not only meet their needs, but also to help them deal with the stress that they felt with examinations:

> Because I think when it was the old style modular system, particularly before they changed it to then say, no modular and doing it all in one sitting, the previous modular one you could sit it in July and sit one in November. That benefited our children because they were able to see 'okay, that's what I need to do in this timeframe and I can have another go'. The linear system, they know it's Year 11 and it's give it your best shot and that's it. It's sort of your one shot. I think what they're finding is that they can't deal with

> pressure and what we're having to do is manage their expectation but also manage parents' expectations. With the linear system the pupils find it hard to be able to revise and I have set for the last four years four different revision packages for pupils, based on sort of their particular needs.
>
> (Head of Mathematics 12)

Mental health

> So yeah, we have also had to do a lot of mental health awareness, mindfulness side of things because that has spiked, that has been something again difficult to deal with where you know, every school has its fair share of mental health issues to deal with on a day-to-day basis. I think then saying 'Okay, by the way you're now having all of these exams and everything's really important', in just this four-week, five-week period that makes people very, very stressed and very, very ... and for the people who don't deal with their emotions well at the best of times and struggle with just what happens on a day-to-day that really does have a cataclysmic effect on some of the students.
>
> (Vice Principal in Charge of Maths 4)

Fifteen teachers from eight different schools explained that with the introduction of the linear examinations particularly, over the last two years, there had been an increased need to focus on students' mental health. Citing the lack of coursework or speaking and listening assessments to absorb some of the pressure (Head of English 2), and that a lot depended on many examinations that students sit over a relatively short period of time (Head Teacher 9), teachers reported introducing a number of interventions to help combat the high levels of stress. A degree of anxiety and stress was also generated by the level of uncertainty created by a lack of clarity over the new grading structure (Head of Maths 11) right up until the start of the new examinations. One teacher said that a knock-on effect of the increased anxiety was that students struggled with their self-esteem (Vice Principal in Charge of Maths 4) and another commented on the impact on their confidence to be able to remember everything that they needed for the terminal examinations (Head of English 4). Three teachers also expressed concern over what they predicted as increased anxiety when all the subjects moved to linear examinations (Head of English 9). The issue of anxiety over the terminal examinations and its effects on students' wellbeing was felt particularly keenly at the special school, where all three interviewees spoke of heightened anxiety amongst their students.

> More pressure. I mean, two out of the eight weren't here this year, they signed off sick.
>
> (Head Teacher 12)

Strategies and interventions

In response to the elevated levels of stress associated with that summer's examinations, teachers cited interventions that had been implemented both at the school and classroom level to help students:

> Yes, it's been stressful. They've generally found it really stressful, and so you can put little things in to try to support, like little meditation groups – we've got those running. We can do little yoga groups; we can do time-out; we can think about doing extra support; we can do revision classes every lunch day for seven weeks, which we did for English, to try to make sure that people had the support that they felt they needed, but ultimately, it's scary doing terminal exams.
>
> (Head of English 9)

When talking about preparation for exams, some teachers spoke of how it was becoming increasingly about mental preparation for pressure (Vice Principal in Charge of Maths 4). Three schools cited putting on extra sessions after school, mapping out things that their students could do every day to help combat stress. Two different schools cited counselling sessions to help students combat their heightened anxiety:

> We do everything in our powers to reduce anxiety, we have an educational psychologist here for three days a week, I don't know of another school in the maintained sector that has that, and I include mainstream schools … everything is geared to reducing anxiety.
>
> (Head Teacher 12)

Progression to A levels

Progress to A level featured quite heavily in the interviews, but in a varied manner when talking about the different groups of students. Some teachers critiqued the new examinations in that they felt like they were designed with very specific student in mind (Head of English 4) –'the A level kind of ready students' (Director of English and Mathematics 8). Teachers also felt that the new qualifications would be good for 'setting students up for A Level and future' studies (Deputy Head of Academics 11). One teacher worried about 'those students who, while being quite competent to have been around the C/B borderline in previous years', were no longer going to be able to access the new examinations due to the level of difficulty (Vice Principal in Charge of Mathematics 4). One teacher explicitly mentioned that the GCSE examinations were not suited for those students who may have been considering vocational pathways:

> Well it suits students who are designed to do any A level, so I think that, you know the, um, the idea of writing for two and a half hours and then the exam is something at A level they should be prepared for and this does prepare them for it. Um, students who are going to do apprenticeships or BTECs or NVQs, um, perhaps this isn't the right path for them.
>
> (Head of English 2)

Some teachers spoke of how the 'current Year 11s were already thinking in a more kind of academic way than some of [the current] Year 13s even' (Head of

English 2), as there was a lot more A level work that had been pushed down into the new GCSE examinations [structure confounded with curricular content]. One teacher pointed out that when it came to A levels, having the harder examinations at GCSE would be a good thing, as they would already be used to the stress and load (Head of Mathematics 5). Five teachers also commented on how the terminal examinations provided students with a confused message in that high stress was associated with the GCSEs, but A levels would be less of a 'slog' due to there still being coursework (Head of English 4). Several of them pointed out that, where in previous years they had always had a huge take-up at A level, this [coming] year they would not be running the course at all, as the GCSEs had turned their students off taking up A level subjects (Head of Mathematics 9).

How, if at all, has the examination route influenced teaching strategies and teaching materials?

It was also important to understand how teachers perceived they had adapted their practices in the classroom to the changes that occurred in the GCSE examinations. This section focuses on understanding what new practices may have been implemented; trying to understand how linearity has influenced the need (or not) for change; and how existing practices may have been adapted.

Assessment culture

Most of the schools interviewed reported a change in the way that they approached assessments in their schools, in part driven by linearity, and in part by the changes to the structure of the questions in the GCSE examinations. These changes have included more examinations, or changes in the way that the schools prepare their students for taking an examination.

Increased assessment and examination skills

In ten of the schools where interviews took place, additional mock examinations or other styles of assessment had been introduced to provide support similar to that given by modular examinations. These additional examinations enabled the teachers to provide feedback to their students (Head of Mathematics 12), provided a sense to students of what sitting examinations would really feel like (Head of English 5), and helped to ease some of the pressure on students. Teachers felt that students needed to get a better feel for what the new examinations would be like; given that students' grades were entirely dependent on the examinations, they needed more formalised practice:

> Um, one of the big differences of course with 100% exam is they need more exam practice and by the nature of these exams as well, it's very specific things that are being sought in your answers, um, so we have end of Year 10 trial exams, which we've always done, but we've formalised it a lot more, so in the past it would've been in the classroom. From this year on it'll be in the hall, so more of an exam experience. They've always done trial exams in December in Year 11, but the big change then is from January through till April they've basically got a trial exam every month.

(Head of English 7)

Not only was there an increase in the number of mock examinations held, there was also a change in the types of skills that the teachers were focusing on to prepare their students for the examinations. The only exceptions to be noted were at the grammar school and one of the independent schools interviewed which reported no changes to the way they prepared their students for the linear examinations.

> We like linearity, because our philosophy, as I say many times to prospective parents, is that we teach the subject, we don't teach to the test, we don't teach to specification. And so, by teaching the subject, the overview of the subject, it's a much better experience for our able pupils, and they can then take the exams in their stride. So, we cover everything, but its linearity allows us to not to break it up, and become excessively exam driven.
>
> (Head of Mathematics 6)

The deputy head teacher at the grammar school commented that whilst examinations may have had a modular structure before, the modular courses over the years had become more 'synoptic' in nature, requiring schools to have a revision or summary period just before the examinations, and so a move to a linear structure should not have proved to be a big jump. For their school, the linear structure of the qualifications was only a 'positive' thing because it encouraged 'students to see it [the curriculum] as a whole'. For the teachers at 10 schools however, the linear nature of the examinations had shifted focus on the types of skills the students needed to develop and hone, in addition to their knowledge of the curriculum. Teachers spoke of providing more practice of 'writing in timed conditions' (Head of English 2), recall exercises, recap lessons, and constantly having to visit and revisit ways in which to do revision.

> I mean it's really also been looking at how we teach revision, so looking, you know, making sure that we're sort of explicitly teaching things like dual coding, so when we're, you know, when the students are revising quotations, very simply drawing silly images or something that's going to help them sort of, you know, remember the quotation in their heads and sort of really explicitly teaching the fact that reading a page of notes isn't revising……the importance of revision had never been so great for these students but they'd really lower down the school had never had to build up any mechanisms for it because we teach a unit, we test it, we teach a unit, we test it and it's and they sort of never had to do it independently in that way.
>
> (Head of English 4)

One teacher explained that their school had put in a second set of mock examinations to help students with their 'lack of confidence [in] what their actual exams are going to be like and how much more content there is in them' (Deputy Head of Academics 11). The special school had introduced mock assessments for the first time formally, due to a 'drive from parents and carers'

(Head of Mathematics 12), stemming from the uncertainty that surrounded the new examinations. At another school, students were now sitting mock examinations in language and in literature split throughout the year, so 'that they have the experience of doing the whole paper' (Head of English 9), which is not something that they had done before at the school. Other schools spoke of how they were now pushing examinations further down the years, and introducing their students to not only the examination structure but also to the skills needed to approach the examinations from Key Stage 3:

> Um, I think more than anything it's preparing students to write in timed conditions in a lot more … um, getting them really used to that, because it's all they will do. So even at Key Stage 3 they're doing GCSE style assessments.
>
> (Head of English 2)

Matching the internal to external assessment

> I feel like we're kind of spoon feeding them a bit and the exams are quite, the questions are quite restrictive, so I don't know if they're going to be those independent thinkers that the qualification was designed to produce.
>
> (Head of English 2)

There were a few teachers who felt that the new linear examination structures were creating less independence for the students, as the students were needing to think about very specific ideas for questions, which wasn't necessarily the 'kind of quality that AQA was hoping or the government was hoping for' (Head of English 2). Some teachers were critical that the structure for the internal examinations and mini-assessments during class time was all focused around how pupils should approach and answer the kinds of questions that they would now see on the examinations (Head of Mathematics 7), and less about teaching the content for the examination (Head of English and Mathematics 8).

On the other hand, some teachers felt that moving away from the cycle of teaching-assessment-teaching-assessment, the 'treadmill' (Head of Mathematics 9) of modularity, was actually making their teaching and the learning better. One teacher even pointed out that modular examinations had thus far trained students into approaching questions with a sense of familiarity, 'an "OK, well I can do this because this lesson is about" [in reference to the exams directly reflecting the questions]' (Vice Principal of Curriculum 4) because there had been less need for the longer term, deeper understanding that they felt came with the linear examinations.

Setting

Whilst the majority of the teachers talked about having an unchanged policy with regard to setting or banding their students (particularly in mathematics), a few explicitly mentioned the changes to the examinations, and in particular the nature of the qualifications, as the reason for changing the way they set or grouped their students in classes. At one school, the removal of tiering in English

had triggered a move to mixed ability classes, and this had filtered down to Key Stage 3 to provide continuity through to Key Stage 4 (Head of English 9). One teacher mentioned that the removal of tiers 'does present that opportunity [of mixed sets] that we haven't previously had' (Head of English 4). At another school, the head of English mentioned that by having a 'tierless qualification …. there's real scope from moving away from' (Head of English 2) setting, which in turn would help with parental complaints about their child feeling stigmatised about being in the bottom set, or would even help that child no longer feel like they were the lowest priority in the school.

In mathematics, where there were still tiers, but the demands of each tier had changed, some teachers talked about having to re-evaluate how many of their students they would enter for the higher tier. For example, where previously one school had had four sets of students working on the higher tier content, this year it had only two, marking a significant shift for the school in its setting strategy (Head of Mathematics 2). At another school, whilst the teacher spoke of setting in the same manner as before (Head of Mathematics 11), the same change in tier entry pattern was seen where nearly double the number of students had been entered for the foundation tier than the previous year.

Curriculum and timetable organisation

Eleven teachers (across mathematics and English) from seven different schools indicated no change to the way their curriculum was organised and delivered at their schools. This was either because their curriculum had always been structured to deliver the linear qualifications (Deputy Head of Academics 5), or because, even if they had wanted more lessons in the week, they had not scheduled them due to the time needed for all the GCSEs that were offered in that school (Head of Mathematics 9), or they felt that they could manage the demands of the new specification without any adjustments to the curriculum structure (Head of English 6):

> The teachers have remained the same, because they've been teaching GCSE English for a long period of time. The timetables have also remained the same, because you change one timetable, the whole thing goes into orbit. So, these two things have remained the same … Again, just to refine your question, my answer, very little changes have been made to the current curriculum way.
>
> (Head of English 1)

Where schools did talk about curriculum reorganisation, various means of doing so were discussed. At the special school for example, the Head of English was not planning to enter all the students for the English literature examination in the coming year due to the increased demands of the qualification. In English, students in Year 10 are no longer able 'to have time off, golden time, for good behaviour, and for working hard' (Head of English 12), as that time is now all allocated to teaching the curriculum. This was perceived as a significant and unfortunate change for the students at the special school.

The time allocated to the core subjects in schools was also varied. Some teachers alluded to having more teaching time as no time was needed for controlled assessments. In some cases, this 'created' time was used to tackle examination-style questions in class (Head of Mathematics 1). In one school, mathematics gained an extra lesson a week in Year 10 to deliver the curriculum topics in larger chunks (Head of Mathematics 3). In fact, 12 interviewees spoke of having more time to teach the new specifications by starting as early as Year 9. This was either through the introduction of certain topics in Year 9 itself (Head of Studies 6), or through the teaching of skills to approach certain questions in the new examinations:

> We've also had to push some of those skills back into Key Stage 3 as well, so the Key Stage 3 curriculum has been completely overhauled as well. So, there will be discrete units on non-fiction that are really looking at question three approaches to non-fiction. There'll be ones that are discretely focussing on summary skills, for example.
>
> (Head of English 5)

At one school, a teacher spoke of adjusting schemes of work all the way back to Year 7, to really build the foundations of skills such as critical analysis and problem solving that would be needed to tackle the diverse types of questions on the longer examinations (Vice Principal in Charge of Mathematics 4).

Some schools spoke of having to adjust their schemes of work in English, due to the demands of the new specifications. At one school, for example, a teacher mentioned that due to the higher 'weighting on the technical element of language on the new course' (Head of English 7), they were adjusting the focus on language down to Key Stage 3 to prepare students. At another school, the curriculum organisation had been changed to reflect the effects of linearisation – akin to being on a carousel, where every topic needed to be revisited once to ensure that their students could handle the closed text approach and memorisation requirements (Head of English 9). Teachers found that if a topic was not covered quickly enough, they wouldn't have time to revisit it, and if they didn't have time to revisit it, students couldn't retain the information needed for the examination at the end of the two years.

Mathematics teachers indicated that there had been no changes to their curriculum organisation that had been driven solely by linearisation. For the most part they suggested that it was the order and content of the curriculum that had led to an increase in the amount of time (if at all) that was set aside for teaching the curriculum:

> I would say that is all down to the curriculum rather than the linear nature of it …. Year 10 and Year 11 used to have three lessons a week. Year 10 have now moved up to four lessons a week. Year 11 have stayed on three. We massively changed the scheme of work so that there is more emphasis on, for example, the proportion where there's more of that in the curriculum.
>
> (Head of Mathematics 7)

At the FE colleges, funding was cited as a barrier to changing the structure of the way the curriculum was delivered to meet the demands of the new specifications. For example, even though the mathematics GCSE was now larger and more demanding, there was no flexibility to spend more hours on it as a large part of their funding was tied to the vocational courses (rather than English or mathematics) which were therefore given more time (Director of English and Mathematics 8). At another college, a teacher felt that their jobs were harder (given the attainment profile of their students) as they had to engage with students for only three hours a week and still keep students motivated and engaged in order to maintain funding for the college:

> … one consequence of that is, say, the learner dropped out of maths, the college would get, well, it depends how long they'd been on the course for, but if they're not doing maths, the college gets no money at all for that learner. Because, that's not proportional, it's all or nothing, which means that some learners are very reluctant, and you have to keep them going, basically. Now, I'm a mathematician, I love the subject, and I want learners to enjoy maths, but some of them don't, and don't want to do it. But, that's another issue, really…We get more reluctant learners, that's the other thing we've got to deal with now.
>
> (Head of Mathematics 1)

Impact on teachers

While we have spent the majority of this chapter discussing the impact of the new examinations of schools and students, it is worth noting that some teachers voiced the issues that they themselves have had to contend with in this policy implementation. One head teacher talked about the anxiety that was felt by teachers themselves who were tasked with delivering the new specifications, who 'just feel a little bit at sea' (Head Teacher 9) when dealing with all the changes at the same time. At an FE college, one teacher spoke of the immense pressure to deliver the more demanding content to those students for whom English was not their first language (Head of English 5). In English, at another FE College, a teacher spoke of how they 'felt so much tax' (Director of English and Mathematics 8) in delivering the new English specification as they wanted the freedom to do the work that was needed for the 'good of the learner' (to bring their skills up to level) but were unable to, as they needed to move forward to cover the increased content. At a third FE college, teachers spoke of the extra demands they felt in trying to deliver the content of the new curriculum in only one year's time to those students who were having to re-sit the new GCSEs (Head of English 3). Teachers also felt that they needed to be careful in how they talked to their students about the examinations and grading, as they themselves were 'sort of stabbing in the dark' (Vice Principal Head of Curriculum 4) and still needed to help maintain calm among the students who were facing pressure at many levels.

> And obviously that, you know, with all good will, you try not to put that stress onto the students but I think naturally, you know, your

> anxiety probably is, you know, transferred across to the students at times.
>
> (Vice Principal Head of Curriculum 4)

Most teachers recognised that with the switch to a linear examination, there had been a change in the way they work, and that came at an additional stress and added workload given that they believed that the specifications were 'rushed through at the last minute' (Deputy Head of Academics 11), particularly as the teachers had yet to build their experience with the new curriculum. The teachers also had to deal with added workload in terms of the increased time in the timetable (Head of English 11) or increased examinations. Some teachers spoke positively about how they help each other with the new specification (Head of English 4). Some spoke about not only sharing resources, but coming together as a department to share teaching strategies to do what's best for their pupils (Head of Mathematics 12):

> I think one positive for us though is that the linearity has ... because now when it comes down to it, the whole course is distilled into four, two hour exams, for the very first time, sort of not just discussing in departments, not for the very first time, but much more so than previously. We're not just discussing what we're teaching in content, we're really teaching ... we're really discussing sort of how we're teaching it and we're working together a lot more to sort of produce less in resources that are sort of, that have come out of these really productive discussions.
>
> (Head of English 4)

Some of the comments (8) related to the way in which the policy had been formulated and implemented. Teachers mentioned the volume of change and that they felt it was rushed, could have been better supported, was difficult to evaluate and that it was politically–driven. The following quotations illustrate their concerns.

> ... if we make five or six changes, massive changes at once, how do you know what's successful or not?
>
> (Head of Mathematics 2)

> Bringing in all of the changes at once was quite difficult to deal with, you know, bringing in changes at A level, bringing in changes at GCSE, because it feels like since 2010 we've been on this constantly change sort of treadmill.
>
> (Head of English 4)

> There's tinkering around the edges and there's a sort of whole scale, sort of start again and see where we end up. But information is a big factor. It's getting the information early in order to prepare teachers to prepare students and I don't think that's necessarily been the case.

(Head of Curriculum 4)

> So, there's an awful lot of changes taking place and the way that the new qualifications, I feel, were rushed in too quickly in order for them, from my perspective, to meet government priorities, which was to see significant changes during the term of the coalition … I feel there was a lack of time to sit back and prepare.
>
> (Deputy Head 11)

As in the baseline research, teachers considered that a period of stability would be welcome, as the following quotation shows:

> I think the worst thing that could happen is there be yet more change … just give it time and be more evidence-based before any further changes are considered
>
> (Headteacher 10)

Discussion of research in schools

The teachers from this phase of interviews were vocal in a belief that linearity, on the system level, had made the examinations seem fairer. Due to this increased sense of fairness and a growing belief that the linear examinations allowed teachers to teach content in a more comprehensive manner, eight teachers identified themselves as having become more pro-linear examinations over the period of the last two years. Whilst providing a more level playing field, by reducing the opportunities to 'game the system', teachers were acutely aware that this fairness was not felt by all students. Teachers were appreciative of the fact that the new GCSEs allowed the higher attaining students to truly shine, however, they were cognisant that the new structure and content were geared towards a route into A levels, and would potentially leave more of the lower attaining students behind. There was also a sense that those students who were more interested in a vocational route would lose out from the increased emphasis on the more traditional curriculum.

Teachers in this second phase of interviews were keen to voice their concerns over the increased stress that their students were feeling due to the new linear qualifications. In the first phase of interviews, these teachers had held mixed opinions about the increased pressure students would feel. These fears, however, were realised by the majority of interviewees in the second phase. There was an emphasis on the mental health issues students were facing and teachers explained that they would have to provide further intervention strategies as all subjects moved towards linear examinations structures.

In the first phase of interviews there was a sense among the teachers interviewed that there would be major curricular changes, and much more time allocated in the timetable for mathematics and English. In the second phase of interviews, however, while teachers talked about feeling as if focus was being directed into the core courses, there were not as many substantial changes in the

timetable in schools as teachers had initially anticipated. This was partly confounded by funding issues that some schools were facing at this time.

Teachers had proved very resilient to the volume of changes that had been made in schools in general, including this switch to linear examinations at both GCSE and A level. Their comments on the nature of the reform process and the need for more information about the reforms stemmed from what they needed to do their jobs well. Teachers were in favour of a period of stability and evidence-based reform.

Chapter 9 Discussion

This century has seen policy shifts in relation to the structure (i.e. linear or modular) of GCSE examinations. A number of claims have been made in favour of each structure, but empirical evidence is lacking and the consequences of implementing linear and modular exams are not well known. Understanding the implications of examination structure policies is critical for policy making, and Ofqual announced its intention to review the impact of the structural reforms of GCSE. In this context, this research has investigated the impact of linear and modular examinations upon outcomes and teachers' reported practices at GCSE. With that, it has aimed to inform evidence-based decision making on the likely effects of future reform agendas in this area.

Evidence presented in this report stems from a research project consisting of a systematic review of the literature, quantitative research, and qualitative research. The systematic review of the literature documented arguments and empirical evidence related to the advantages and disadvantages of modular and linear examinations. The quantitative research strand examined with large-scale datasets the impact of linear and modular exams on GCSE outcomes, changes in the outcome, and GCSE attainment gaps (e.g. SES and gender). The qualitative research strand used semi-structured interviews to assess teacher beliefs on the impact of examination structures and potential influences on teaching practices. A number of findings relevant for policy makers emerge from this research.

Classification schemes for modular assessment structures (Ertl & Hayward, 2010) were useful in thinking about the kind of examination that modular GCSEs had been. The qualification was subdivided into units, where credit for these units was typically only available within the qualification and the curriculum was specific to the qualification: this is an 'internal' approach. Our review of modular assessment showed that it was uncommon for it to be used for high-stakes academic examinations, with some exceptions. Notable amongst these is the entire Grade Point Average (GPA) system of teacher assessment used in the US, which might be thought of as modular and Scotland has a modular assessment regime. School-leaving examinations were made modular very late in England in comparison with vocational qualifications and higher education.

Previous research has generally used critical evaluation rather than empirical methods to draw conclusions about the effects of linear and modular examination structures. Where empirical research was conducted, there was rarely a recognition of the limitations of the data. Overall, the literature review points to claims that linear examinations favour longer term retention of information and foster depth of learning, whereas modular examinations allow a testing-when-ready approach and mastery of topics before moving on to the next area of the course. It is claimed that modular exams can result in higher grades through re-sits, but modularisation is also associated with 'grade inflation', instrumental attitudes to learning and disruptive teaching.

In policy terms, it is remarkable that all of the GCSEs became modular in the 2007 reforms. No explicit policy directive for modular GCSEs was given by the

government or any of their agencies. This was linked with the prevailing policy context and narrative of that time, which stemmed from a drive to integrate vocational and academic qualifications through frameworks.

Standards

Modular examinations proved challenging for the then-current systems of awarding grades and maintaining standards. A combination of re-sitting, banked marks, low weighting of final assessments and early and multiple entries brought about very complex sub-cohorts of students taking the examinations. This made a ubiquitous definition of the correct standard difficult. GCSE science and GCSE English standards in particular proved contentious. One advantage, then, of a linear examination system is that standard setting is more straightforward. The standard setting process is likely to be fairer for linear examinations because differences between examination series can be tackled using conventional methods.

A wealth of research has previously been published on the effects of re-sitting, showing that there were significant effects upon the proportion of students attaining grades, even if individual students typically re-sat a module only once or twice for a qualification. Whether or not the increase in outcomes across time is interpreted as grade inflation, a different question, regarding whether higher outcomes were associated with examination structure, was also of interest. After all, the GCSE standard setting processes were designed to align standards, no matter the structure.

The quantitative analysis with the National Pupil Database and aggregated examination board data focused on three subjects: English, mathematics, and science. The following research questions were addressed:

What are the effects of the examination route on:

- overall GCSE outcomes and changes in the outcome?
- the gender gap?
- the SES (socio-economic status) gap?
- differences between state and independent schools?

Altogether, results of the quantitative strand suggest that GCSE and A level results are unlikely to have been unduly influenced by the examination structure at GCSE. Further, there was no evidence for differential effects of the GCSE examination structure by gender or SES. That is, findings do not support claims that modular or linear exams tend to favour male or female students or influence low and high SES students differently. As expected, high SES students performed better than low SES students in GCSEs and female students performed better than male students in GCSE English. However, GCSE gaps related to gender and SES remained relatively unchanged for linear and modular examinations. Results relating to GCSE gaps between state and independent schools were less clear. There was weak evidence that state schools performed better in modular than linear exams compared to independent schools. It is, however, possible that

entry patterns between state and independent schools are confounded with specification choice rather than modularity or linearity. A level attainment in the same subject did not appear to be affected by having taken a prior modular or linear GCSE.

Economics

England's GCSE examination market is unusual for its structure, as a regulated oligopoly (Jones, 2011). Examination boards invest at times of qualification reform and harvest income when the specifications are in a steady state. Many subjects have to be subsidised by income from the main entry subjects and examination boards share the offer of low entry subjects so that schools can be offered a wide range of curricula. Between 2015 and 2018 examination prices rose by slightly more than the retail price index. This is likely to be due to the costs of reform. Costs of re-sitting were likely to have been significant for schools and therefore the move to the linear structure is likely to have reduced the costs of examinations to the system as a whole. Overall, reform itself is of course a cost, with resources being deployed at all levels of the system to ensure that policies, computer software, administrative guidelines, examination papers, textbooks and so on are updated.

Classroom effects

The qualitative strand relied on interviews with teachers to address the following research questions: What do teachers believe the effects of examination route are upon overall GCSE outcomes and grade inflation? What do teachers believe the effects of examination route are for different groups of pupils? How, if at all, has the examination route influenced teaching strategies and teaching materials according to the teachers? Teachers were interviewed twice: once, before and once after the introduction of linear examinations, in order to investigate the consequences of examination structure reforms on teaching and learning in the classroom from the teachers' points of view.

In terms of effects on overall GCSE outcomes and grade inflation, teachers believed that reforms were designed to combat perceived grade inflation and to reduce rising GCSE outcomes. There were concerns in mathematics that curriculum demands and increased curriculum could cause difficulties for curriculum coverage in timetabled slots. Importantly, however, it was not possible to disentangle in interviews perceptions of teachers related to changes in the examination structure with reforms involving increased curricular demand and more challenging examinations. That is, these perceptions may be confounded.

With regard to differential effects for different groups of pupils, some teachers considered linear examinations more suitable for older pupils and modular examinations more suitable for weaker students, especially if they had poor memory for the subject. Particularly in mathematics, linear examinations were thought to suit stronger students more. Further, teachers somewhat agreed with the stereotype that modular exams suited female students and linear examination suited male students. These perceptions were, however, not

supported by results of the quantitative analysis showing no differential effects of the examination structure by gender or SES.

Teachers deemed modular exams to have motivated students who had done well on early modules, but for the group of students who did poorly on these, it was observed to have been demotivating. In the second round of interviews, teachers more explicitly considered new linear GCSE examinations to favour students doing A levels rather than those who may be more interested in a vocational route. There was significant concern about the rising impact of the linear structure on students' mental health and some teachers indicated that interventions had been introduced to help students combat stress.

In terms of influences on teaching practices, the introduction of linear examinations, somewhat counterintuitively, appears to have led to more school assessments, particularly more mock examinations. There was a shift to teaching revision and examination skills to a greater extent, as the stakes in the linear assessments were higher. The uncertainty created by the reforms, particularly changes to the grading scale, may have partly driven the introduction of more school assessments. This may reduce as the examinations and grading scale bed in.

The shift to linear examinations was perceived positively in terms of allowing for more teaching time with fewer examinations scheduled. Following the implementation of linear examinations, teachers thought examination questions would be less easy to predict and indicated that they would be teaching different skills, such as more problem solving.

Teachers' anxieties about the changes and their potential impact on outcomes were prominent in both rounds of interviews. Yet evidence from the first (2017) and second (2018) awards of the reformed GCSEs showed that results within schools were remarkably stable from year to year. It might be expected that some schools would respond better to the changes than others, creating volatility in results. However, analyses of center level variability showed the converse (Ofqual, 2018) and that schools had responded well to the changes.

Limitations

The qualitative research strand was constrained by several limitations. One is that for schools that had a predominantly linear policy in the first interview round, there was, understandably, less perception of a change to their practices with the examination reforms. Another is related to representativeness of the interview data. A relatively large number of interviews was conducted (49 in the first wave and 35 in the second wave) for a qualitative study, however, the number of teachers is a very small fraction of the population of teachers involved in GCSE teaching in England.

Further, as the baseline wave of research was conducted after the cessation of modular GCSE examinations, it relies upon teachers' memories of their practices prior to the reform. Memory is imperfect, and perceptions of teachers may be confounded with various other reforms to GCSE examinations, though it is likely

that teachers will have been interviewed during a period in which they were still cognisant of the kinds of changes they were making to their practice due to the reforms.

There are other limitations related to the quantitative research. The most important one is that findings are based on observational large-scale data rather than on data collected under experimental conditions. However, since the research relates to the population rather than an experimental sample, this also has advantages. The large dataset is unique in that examination structure data was collected directly from examination boards and matched with NPD data. Regression analysis and matching techniques properly controlled for observed factors related to modular and linear examination structures. However, a number of unobserved factors in the data may be confounded with GCSE performance for linear and modular examinations, such as school choice, the influence of teachers, specification availability, and more subtle, student-level influences. To the extent that students are not randomly assigned to linear and modular examination conditions, it is not possible to rule out other plausible explanations for the effect of the modular and linear structures in the analysis with these data.

Another limitation is that modular and linear assessment routes were not available in each subject in every academic year. Therefore, the outcomes for modular and linear assessment might be confounded with the standard setting process. Yet another limitation is that modular examinations allow unit re-sits before certification. This is not captured in the NPD and therefore the analysis reflects GCSE outcomes to a greater extent than the effects of modular and linear teaching. In terms of data quality, the analysis is limited by missing values likely not missing at random and proxy measures, for example, for capturing the students' SES.

Future research

Little is known about the effects of examination structure upon students' study patterns or their short- and long-term memories for curriculum materials. Randomised control trials on these topics would be difficult to conduct, but would provide very useful information. Pilot examinations would be necessary to ensure that the high-stakes nature of the examinations was retained, as this is likely to have a large impact in itself upon students' practices, effort and motivation. Some pilot examinations have involved students being awarded graded qualifications. Although costlier, it is clear that the quality of data for this project would have been improved had it also encompassed analyses of documents and other data on curriculum, timetabling, examination results and observation of teaching practices. This would have added to the depth of knowledge available on classroom practices. At a systems level, it would have been beneficial to know the economic impact of the removal of modular examinations upon awarding bodies' finances. This is commercially sensitive information, so such research comes with difficulties, but could be conducted, for example, by an insider researcher.

Conclusions

Internationally, modular secondary school examinations are not commonplace. Further, it was a permissive environment rather than proactive policy choice which led to the introduction of England's suite of modular GCSEs in 2007. Policy makers may wish to reflect on the contexts in which central control of key qualification design decisions are prudent, and those contexts where the market is best placed to determine design.

We found no educationally significant effect of examination structure on grade outcomes in English, mathematics or science. Nor did we find any statistical evidence of effects of structure upon progression to A level. So there was no evidence to suggest that either modular or linear GCSEs lead to better educational outcomes. Importantly, equity gaps in gender, socioeconomic status and school type were not affected by structure once prior attainment was taken into account. While some teachers believed that certain groups of students would perform better in a modular system, this view was not supported by analysis of outcome data, where such beliefs could be investigated.

Once teaching was well underway, many teachers considered that linear GCSEs provided more valid assessments of students' performances than did modular examinations. However, concerns about students' mental health were raised, with linear examinations considered to have had a negative impact upon wellbeing for some students. This suggests that well targeted support for students who suffer from examination anxiety is needed.

This may be especially important as we found that more internal, school examinations and more examination preparation had been introduced as a result of the change from modular to linear GCSEs. It may be that the internal assessments will reduce as teachers become more familiar and confident with the new qualifications. While the treadmill of formal examinations has reduced, the corresponding increase in mock examinations means that assessment may still heavily feature in some students' educational experiences.

Increases in grade outcomes (at the time of reform suspected to be unwarranted by policy makers) was tackled by the introduction by Ofqual of the comparable outcomes methodology. Therefore, the change to examination structure did not impact upon this directly. However, there is evidence that the maintenance of standards is more straightforward for linear GCSEs and there have been instances where fairness and public confidence has been undermined by difficulties in grading modular GCSEs. Where qualifications are high stakes for students and teachers, a linear approach is more likely to facilitate comparable standards over time and between exam boards.

Although linear GCSEs are less expensive to the education system than modular GCSEs, the reforms had a significant cost. Teachers adapted to the changed qualifications in a range of ways; there was a lot of upheaval in the system. Indeed, teachers felt that a moratorium on reforms to general qualifications

would be welcome. As such, qualification reform must produce considerable benefits for it to be worthwhile.

Finally, it is important to note that these findings relate only to GCSE. The advantages and disadvantages of modularity versus linearity are likely to vary with the purpose and educational context of any qualification. Better understanding the impact of different design choices on the maintenance of standards, outcomes and teaching and learning is a crucial area of further research.

References

ACME (2011). Early and Multiple Entry to GCSE Mathematics. Available at: http://www.acme-uk.org/media/7392/early%20and%20multiple%20entry%20to%20gcse%20final.pdf

Acquah, D. (2013). *School Accountability in England: Past, Present and Future*. CERP.

Adelman, C. (2008). The Bologna Club. What U.S. Higher Education can learn from a decade of European reconstruction. *Institute for Higher Education Policy*, Washington DC.

Alldritt, C., & Taylor, R. (2014). Lack of options: how a pupil's academic choices are affected by where they live. *RSA: Open Public Services Network*.

Alpha Plus (2012). The evaluation of the impact of changes to A levels and GCSEs. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/197437/DFE-RB203_1_.pdf

Anderson, S., Auquier, A., Hauck, W. W., Oakes, D., Vandaele, W., & Weisberg, H. I. (1980). *Statistical Methods for Comparative Studies: Techniques for Bias Reduction*. Wiley Series in Probability and Mathematical Statistics.

AQA (2012a). *AQA GCSE English and English language January 2011 to June 2012: Report to Ofqual*. Manchester, AQA.

AQA (2012b). *What is the impact of re-sitting at A level?* Available at: https://cerp.aqa.org.uk/informing-policy/A level-reform

Arendasy, M., & Somer, M. (2013). Quantitative differences in retest effects across different methods used to construct alternate test forms. *Intelligence*, 41(3), 181-192.

Association of Colleges (2014). The Department for Education Budget after 2015. Available at: https://www.aoc.co.uk/sites/default/files/The%20Department%20for%20Education%20budget%20after%202015.pdf

Atkins, M. J., Beattie, J. & Dockrell, W. B. (1993). *Assessment Issues In Higher Education*. Newcastle upon Tyne: School of Education, University of Newcastle upon Tyne.

Bachan, R. (2015). Grade inflation in UK higher education, *Studies in Higher Education*, DOI: 10.1080/03075079.2015.1019450

Baird, J. (2000). Are examination standards all in the head? Experiments with examiners' judgments of standards in A level examinations. *Research in Education*, 64, 91–100.

Baird, J. (2007). Alternative conceptions of comparability. In: P. E. Newton, J. Baird, H. Goldstein, H. Patrick, & P. Tymms (Eds.) *Techniques for monitoring the comparability of examination standards*, pp.124–56. London, Qualifications and Curriculum Authority.

Baird, J. & Elliott, V. (2018). Metrics in education – control and corruption. *Oxford Review of Education*, 44, 5, 533 – 544.

Baird, J. & Hopfenbeck, T. N. (2016). Curriculum in the Twenty-First Century and the Future of Examinations (Chapter 51) In: D. Wyse, L. Hayward, & J. Pandya (Eds.) *The Sage Handbook of Curriculum, Pedagogy and Assessment*, Vol. 2.

Baird, J., & Scharaschkin, A. (2002). Is the whole worth more than the sum of the parts? Studies of examiners' grading of individual papers and candidates' whole A- level examination performances. *Educational Studies*, 28, 143–162.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. Available at: https://doi.org/10.18637/jss.v067.i01

Bekhradnia, B. (2004). Credit Accumulation and Transfer, And The Bologna Process: An Overview. *Higher Education Policy Institute*, 2004. Available at: http://www.hepi.ac.uk/wp-content/uploads/2014/02/13CAT-ExecSummary.pdf

Betts, M. & Smith, R. (1998). *Developing the Credit Based Modular Curriculum in Higher Education*. London, Falmer Press.

Black, P., Harrison, C., Osborne, J., & Duschl, R. (2004). *Assessment of Science Learning 14-19*. London, Royal Society.

Bolker, B. (2017). *GLMM FAQ: Inference and confidence intervals*. Available at: http://bbolker.github.io/mixedmodels-misc/glmmFAQ.html#inference-and-confidence-intervals.

Brown, S. & Saunders, D. (1995). The challenges of modularization. *Innovations in Education and Training International,* 32, 96-105.

Cadwallader, S., & Tremain, K. (2013). *How policy formation and implementation interacts with risks to high stakes qualifications*. Manchester, AQA Centre for Education Research and Practice.

Clarke, J. (1996). Modularising A levels in the Social Sciences: some conclusions. *Social Science Teacher*, 26(1), 33-35.

Cochran, W. G., & Rubin, D. B. (1973). Controlling Bias in Observational Studies: A Review. *Sankhyā: The Indian Journal of Statistics, Series A*, 35(4), 417–446. Available at: https://doi.org/http://www.jstor.org/stable/25049893

Coe, R. (2004). Issues arising from the use of effect sizes in analysing and reporting research. In: I. Shagen & K. Elliot (Eds.) *But what does it mean? The use of effect sizes in educational research*, pp. 80–100.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. London: Academic Press.

Cresswell, M. J. (1996). Defining, setting and maintaining standards in curriculum-embedded examinations: Judgmental and statistical approaches. In: H. Goldstein & T. Lewis (Eds.) *Assessment: Problems, developments and statistical issues*, pp. 57-84. Chichester, John Wiley and Sons.

Daugherty, R. (1995). *National Curriculum Assessment: A Review of Policy, 1987-1994*. Psychology Press.

Davies, D. (1991). Modular method. *Education*, 177(21), 420-421.

de Waal, A. (2009). *Straight As? A level teachers' views on today's A levels.*London: Civitas. Available at: http://www.civitas.org.uk/pdf/straight_a's.pdf

Dearing, Sir Ron (1996). *Review of Qualifications for 16 – 19 year olds*. SCAA. Available at: http://www.educationengland.org.uk/documents/dearing1996/dearing1996.html

Dearing, Sir Ron (1997). *Higher Education in the learning society*. London: Her Majesty's Stationery Office. Available at: http://www.educationengland.org.uk/documents/dearing1997/dearing1997.html

Department for Education (DfE) (2010). *Importance of Teaching: Schools White Paper*. London, Her Majesty's Stationery Office.

DfE (2011). *End for GCSE modules and spelling, punctuation and grammar marks restored to exams*. Press release, 14 December 2011.

DfE (2013). *Multiple entry to GCSEs: Memorandum to the Education Select Committee.* London, DfE.

DfE (2014). Key Stage 4 Programme of Study for Science. Available at: https://www.gov.uk/government/consultations/national-curriculum-reform-england-key-stage-4-science

DfE (2016). (*Early Entry Guidance for Schools.* Available at: http://www.education.gov.uk/schools/performance/secondary_14/GCSE_Early_Entry_Guidance.pdf

DfE (2017). Supporting mental health in schools and colleges. NatCEN Social Research & the National Children's Bureau Research and Policy Team. August. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/634725/Supporting_Mental-Health_synthesis_report.pdf

DfE (2018). *Statistics: GCSEs (key stage 4).* Available at: https://www.gov.uk/government/collections/statistics-gcses-key-stage-4

DfES (2001). *Special Educational Needs Code of Practice*. Available at: https://doi.org/DfES/581/2001

Department for Education and Skills (DfES) (2005). *14-19 Education and Skills*. February. Available at: http://webarchive.nationalarchives.gov.uk/20050301194752/http://www.dfes.gov.uk/14-19/documents/Final%20Report.pdf

DfES (2013). GCSEs in vocational subjects A general guide and overview of the qualifications. Available at: http://webarchive.nationalarchives.gov.uk/20130401151715/http://www.education.gov.uk/publications/eOrderingDownload/DfES0809200MIG1065.pdf

Donnelly, J., & Ryder, J. (2011). The pursuit of humanity: curriculum change in English school science. *History of Education*, 40(3), 291-313.

Ertl, H. (2001). *The role of EU programmes and approaches to modularization in vocational education: fragmentation or integration?* Dissertation. ISBN 3-8316-0182-8.

Ertl, H., & Hayward, G. (2010). Modularization in Vocational Education and Training. In: P. Peterson, E. Baker, B. McGaw (Eds.) *International Encyclopedia of Education*. Volume 8, pp. 383-390. Oxford, Elsevier.

Forster, M. (2005). *Can Examiners Successfully Distinguish Between Scripts that Vary by only a Small Range on Marks*? Unpublished internal paper. Cambridge: Oxford Cambridge and RSA Examinations.

Frontier Economics (2015). *Understanding awarding organisations' commercial behaviour before and after the GCSE and A level reforms.* Ofqual/15/5596. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/407337/2015-02-26-frontier-report-understanding-ao-commercial-behaviour-before-and-after-the-gcse-and-A level-reforms.pdf

Gill, T. (2013). Early entry GCSE candidates: Do they perform to their potential? *Research Matters*. Cambridge, Cambridge Assessment.

Gill, T., & Suto, I. (2012). *Students' and teachers' views and experiences of A level module re-sits.* Cambridge Assessment Paper presented at British Educational Research Association conference, Manchester, September 2012.

Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning Variation in Multilevel Models. *Understanding Statistics*, 1(4), 223-231.

Gray, E. A. (2002). *The comparability between modular and non-modular examinations at GCE Advanced level*. PhD thesis. UCL Institute of Education.

Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*, 92, 373-385.

Hayward, G., & McNicholl, J. (2007). Modular mayhem? A case study of the development of the A-level science curriculum in England. A*ssessment in Education: Principles, Policy & Practice*, 14(3), 335-351.

Heinrich, M. & Stringer, N. (2012). The Effects on Schools and Pupils of Modularising GCSEs. CERP.

Higton, J., Noble, J., Pope, S., Boal, N., Ginnis, S., Donaldson, R., & Greevy, H. (2012). *Fit for Purpose? The view of the higher education sector, teachers and employers on the suitability of A levels*. Available at: https://www.gov.uk/government/publications/are-A levels-fit-for-purpose-views-of-those-who-use-them

Hirst, K., & Meacock, S. (1999). Modular A levels and undergraduates. *Teaching Mathematics and its Applications,* 18(3), 122-127.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3), 199-236. Available at: https://doi.org/10.1093/pan/mpl013

Hodgson, A., & Spours, K. (2003). *Beyond A levels: Curriculum 2000 and the Reform of 14-19 Qualifications*. Routledge, London.

Hodgson, A., & Spours, K. (2005). The learner experience of Curriculum 2000: implications for the reform of 14–19 education in England, *Journal of Education Policy*, 20(1),101-118.

Hodgson, A., Spours, K., & Waring, M. (2005). Higher education, Curriculum 2000 and the future reform of 14–19 qualifications in England, *Oxford Review of Education*, 31(4), 479-495.

Holding, G. (1995). Implementing Modular A Levels. London, Further Education Development Agency (FEDA).

Horner, M., & Ryder, J. (2014). *Measuring the impact of GCSE Twenty First Century Science on progression to post-compulsory science courses*. London, Nuffield Foundation.

House of Commons Education Select Committee (2012). *Education Committee - First Report. The administration of examinations for 15-19 year olds in England*. Available at: http://www.publications.parliament.uk/pa/cm201213/cmselect/cmeduc/141/14102.htm

House of Commons Education Select Committee (2013a). *Memorandum to the Education Select Committee on Multiple Entry to GCSEs July 2013*. Available at: www.parliament.uk/

House of Commons Education Committee (2013b). *From GCSEs to EBCs: the Government's proposals for reform Eighth Report of Session 2012–13*. London, Her Majesty's Stationery Office.

Howieson, C., Raffe, D., Spours, K. & Young, M. (1997). Unifying Academic and Vocational Learning: The State of the Debate. *Journal of Education and Work*, 10(1), 5-35.

Isaacs, T. (2010). Educational Assessment England. *Assessment in Education: Principles, Policy & Practice*, 17(3), 315-334.

Isaacs, T. (2012). Assessment in education in England SA-eDUC 9 (1). Available at: http://www.nwu.ac.za/node/16728

Isaacs, T (2013a). What does good upper secondary education for the tail – and maybe everyone else – look like? In: P. Marshall (Ed.) *The Tail: how England's schools fail one child in five – and what can be done.* London, Profile Books, pp. 111-124.

Isaacs, T. (2013b). The Diploma qualification in England: an avoidable failure? *Journal of Vocational Education and Training,* 65(2), 277-290.

Isaacs, T. (2014). Curriculum and assessment reform gone wrong: the perfect storm of GCSE English. *Curriculum Journal*, 25(1), 130-147.

Joint Council for Qualifications (JCQ) (2012). News release: Rise in take up of languages and separate sciences as overall top grades dip, pp. 1–2. Available at: https://www.jcq.org.uk/media-centre/news-releases/news-release-for-gcse-and-entry-level-certificate-results-summer-2012

JCQ (2014). *GCSE and Entry Level Certificate Results Summer 2014. From GCSEs to EBCs: the Government's proposals for reform Eighth Report of Session 2012–13*. Available at: http://www.jcq.org.uk/examination-results/gcses/gcse-and-entry-level-certificate-results-summer-2014

Jones, B. (2011). *Regulation and the qualifications market*. Available at: https://cerp.aqa.org.uk/research-library/regulation-and-qualifications-market

Kulik, J. A., Bangert-Drowns, R. L., & Kulik, C-L. (1984). Effectiveness of coaching for aptitude tests. *Psychological Bulletin*, 95(2), 179-188.

Lambert, D. & Lines, D. (2000). *Understanding assessment: purposes, perceptions, practice*. London, Routledge.

Long, R. (2015). GCSE, *AS and A level reform*: briefing paper #06962. London, House of Commons Library.

MacLeod, D. (2007). No limit on A level re-sits, watchdog rules. *The Guardian,* 21 September. Available at: http://www.guardian.co.uk/education/2007/sep/21/schools.alevels.

Mansell, W. (2007). *Education by Numbers: The Tyranny of Testing*. London, Portico's Publishing Ltd.

Mansell, W., & James, M. (2009). *Assessment in Schools: Fit for purpose? A Commentary by the Teaching and Learning Research Programme.* London, Economic and Social Research Council. Teaching and Learning Research Programme.

McClune, B. (2001). Modular A levels – who are the winners and losers? A comparison of lower-sixth and upper-sixth students' performance in linear and modular A level physics. examinations, *Educational Research*, 43(1), 79-89.

McVittie, J. (2008). *SQA Research Report #3: National Qualifications: A Short History.* Available at: www.sqa.org.uk

Meyer, L. (2011). *The Value of GCSEs*. Available at: https://cerp.aqa.org.uk/sites/default/.../CERP-RP-LM-01062011_0.pdf

Millar, R. (2011). Reviewing the National Curriculum for science: opportunities and challenges. *The Curriculum Journal*, 22(2), 167-185.

Millar, R. (2012). *Rethinking Science Education for All*. York, Department of Education, University of York.

Moon, B. (Ed.) (1988). *Modular Curriculum*. London, Paul Chapman.

National Audit Office (NAO) (2016). *Department for Education. Financial sustainability of schools. Report by the Comptroller and Auditor General*. HC 850 Session 2016 – 17. 14 December. Available at: https://www.nao.org.uk/wp-content/uploads/2016/12/Financial-sustainability-of-schools.pdf

NAO (2017). Department for Education. *Retaining and developing the teaching workforce*. HC 307 Session 2017 – 19. 12 September. Available at: https://www.nao.org.uk/wp-content/uploads/2017/09/Retaining-and-developing-the-teaching-workforce.pdf

NAO (2018a). *Department for Education. Delivering STEM (science, technology, engineering and mathematics) skills for the economy*. 17 January. Available at: https://www.nao.org.uk/wp-content/uploads/2018/01/Delivering-STEM-Science-technology-engineering-and-mathematics-skills-for-the-economy.pdf

NAO (2018b). *Department for Education. Converting maintained schools to academies*. 22 February. Available at: https://www.nao.org.uk/wp-content/uploads/2018/02/Converting-maintained-schools-to-academies.pdf

Ofqual (2009a). *GCE, GCSE, Adva*nced *Extension Award Code of Practice*. Available at: http://www.ofqual.gov.uk/61.aspx

Ofqual. (2009b). *The New GCSE Science examinations: findings from the monitoring of the new GCSE science specifications; 2007 to 2008*. Coventry, Ofqual.

Ofqual. (2010). *Review of standards in summer 2009 GCSE science and GCSE additional science*. Coventry, Ofqual.

Ofqual (2011). *GCSE, GCE, Principal Learning and Project Code of Practice*. May. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/371268/2011-05-27-code-of-practice.pdf

Ofqual (2012). *Effects of Unitisation in 2009 GCSE Assessments Comparison of Candidate Achievement in Modular and Linear Assessments*.

Ofqual (2013a). *Annex C. Early entry in summer 2012 and summer 2013 GCSEs and IGCSEs*. Ofqual/13/5313. Available at: http://webarchive.nationalarchives.gov.uk/20140813095715/http:/ofqual.gov.uk/news/ofqual-has-written-to-the-education-select-committee/

Ofqual (2013b). *Reforms to GCSEs in England from 2015: Regulatory Impact Assessment*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/529389/2013-11-01-regulatory-impact-assessment-on-reforms-to-gcses-from.pdf

Ofqual (2014) *Setting GCSE, AS and A Level Grade Standards in Summer 2014 and 2015.* Ofqual/15/5759. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/451321/2015-08-05-summer-series-gcse-as-and-A level-grade-standards.pdf

Ofqual (2015a). *School and College Purchasing Behaviours*. February. Ofqual/15/5594.

Ofqual (2015b). *Further Decisions for Completing GCSE, AS and A Level Reform in 2017*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/479635/2015-05-14-reform-of-gcses-as-and-A levels-in-2017-may-2015.pdf

Ofqual (2015c). *Assessment of Practical Work in New Science GCSEs – Summary*. Available at: www.ofqual.gov.uk

Ofqual (2015d). *GCSE Science: Consultation on Conditions and Guidance*. Available at: www.ofqual.gov.uk

Ofqual (2017a). *Entries and late entries for GCSE and A level: 2016 to 2017 academic year. Background information accompanying the statistical release*. 23 November. Ofqual/17/6307.2. Retrieved from: https://www.gov.uk/government/statistics/entries-and-late-entries-for-gcse-and-A level-2016-to-2017-academic-year

Ofqual (2017b). *Maintaining standards in the first awards of the 9 to 1 GCSEs Principles and evidence to inform the approach for summer 2017.* Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/667503/6328_-_Maintaining_standards_in_the_first_awards_of_the_9_to_1_GCSEs.pdf

Ofqual (2018). *Variability in GCSE results for schools and colleges 2016 to 2018*. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/735661/Variability_in_GCSE_Results_for_Schools_and_Colleges_2016_to_2018.pdf

Ofsted (2013). *Schools' use of early entry to GCSE examinations.*

Pearson (2011). *GCSE Science Specifications*. Available at: http://qualifications.pearson.com/en/qualifications/edexcel-gcses/science-2011-mixed.html

Pilz, M. (2002). The contrast between modular and occupational approaches to modernizing vocational training. *Vocational Training: European Journal*, No.25. 27-33.

Priestley, M. (2003). Curriculum 2000: A broader view of A levels? *Cambridge Journal of Education*, 33(2), 237-255.

Priestley, M. & Minty, S. (2013). Curriculum for Excellence: 'A brilliant idea, but...' *Scottish Educational Review,* 45 (1), 39-52.

QCA (1998). *Qualifying for Success: report on the consultation about the future of post-16 qualifications*. London, QCA.

QCA (2004). *GCSE, GCSE in vocational subjects, GCE, VCE, GNVQ and AEA Code of Practice 2004/5.* London, QCA.

QCA (2005). *A review of GCE and GCSE coursework arrangements*. London, QCA.

QCA (2007a). *GCSE qualification criteria*. London, QCA.

QCA (2007b). *GCSE qualification and subject criteria consultation summary report*. QCA/07/3473. London, QCA

QCA (2007c). *A level re-sitting: su*mmary of research findings. Available at: http://dera.ioe.ac.uk/id/eprint/8669

QCA (2008). *Regulatory arrangements for the Qualifications and Credit Framework.* Crown Copyright. ISBN 978-1-84721-674-8. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/371294/2008-08-15-regulatory-arrangements-qcf-august08.pdf

Raffe, D. (1994). Modular strategies for overcoming academic/vocational divisions: issues arising from the Scottish experience. *Journal of Education Policy* 9(2), 141-154.

Raffe, D. (2009). The Action Plan, Scotland and the making of the modern education. *Scottish Educational Review*, 41(1), 22-35.

Raffe, D. (2011). Are "communications frameworks" more successful? Policy learning from the Scottish Credit and Qualifications Framework. *Journal of Education and Work*, 24, (3–4), 283–302.

Richardson, W., & Sing, S. (2011). *The impact of practical and 'vocational' learning on academically-able young people aged 11–16.* A report for the Edge foundation. London, Edge Foundation. Available at: http://www.edge.co.uk/media/76167/exeter_academically_able_report_oct_2011.pdf

Robbins (1963). *Committee on Higher Education. Higher education: report of the Committee appointed by the Prime Minister under the Chairmanship of Lord Robbins 1961-63*, Cmnd. 2154, London, HMSO.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.

Rosenthal, J. A. (1996). Qualitative Descriptors of Strength of Association and Effect Size. *Journal of Social Service Research*, 21(4), 37-59.

Ryder, J. & Banner. I. (2011). Multiple Aims in the Development of a Major Reform of the National Curriculum for Science in England. *International Journal of Science Education*, 33(5), 709-725.

Ryder, J. & Banner, I. (2013). School Teachers' Experiences of Science Curriculum Reform. *International Journal of Science Education*, 35(3), 490-514.

Sävje, F., Higgins, M. J., & Sekhon, J. S. (2016). *Generalized Full Matching*.

Sävje, F., Sekhon, J., & Higgins, M. (2017). Quick Generalized Full Matching: quickmatch. Available at: https://github.com/fsavje/quickmatch

Scharaschkin, A., & Baird, J. (2000) The effects of consistency of performance on A level examiners' judgments of standards. *British Educational Research Journal*, 26, 343–357.

Scott, E.P. (2012). Short-term gain at long-term cost? How resit policy can affect student learning. *Assessment in Education: Principles, Policy & Practice*, 19(4), 431-449.

Slowey, M. (1995). *Implementing Change from Within Universities and Colleges: 10 Personal Accounts.* Kogan Page.

Smith, A. (2004). *Making mathematics count: The report of Professor Adrian Smith's inquiry into post-14 mathematics education*.

Smith, G., & Smith, J. (2005). Regression to the Mean in Average Test Scores. *Educational Assessment,* 10(4), 377-399, DOI: 10.1207/s15326977ea1004_4.

Smith, R., & Bradley, G (1996). The Americanisation of British higher education. *Reflections on Higher Education*, 8, 40-55.

Snijders, T., & Bosker, R. (1994). Modeled variance in two-level models. *Sociological Methods & Research*, 22(3), 342-363.

Sperring, R., Davey, J., Jones, K. & Anderson, T. (2017). *Approaches to early and multiple entry for GCSE examinations in Wales.* Qualifications Wales. Available at: https://www.qualificationswales.org/media/2825/approaches-to-early-and-multiple-entry-2017-e.pdf

Sutch, T. & Wilson, F. (2013). Reaching for the A*: Exploring the extent and effect of re-sitting at A level. *Research Matters*. 4–48. Cambridge, Cambridge Assessment.

SQA (2013). *Assessment in the new National Courses and Units*. Glasgow, Scottish Qualifications Authority. Available at www.sqa.org

Strand, S. (2004). The use of effect sizes: two examples from recent educational research. In: I. Schagen, K. Elliott (Eds.) *But what does it mean? The use of effect sizes in educational research,* pp. 42–54.

Strand, S. (2006). Comparing the Predictive Validity of Reasoning Tests and National End of Key Stage 2 Tests: Which Tests are the Best? *British Educational Research Journal*, 32(2), 209-225.

Stuart, E. A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science*, 25(1), 1–21. Available at: https://doi.org/10.1214/09-STS313.

Taverner, S. & Wright, M. (1997) Why go modular? A review of modular A level Mathematics. *Educational Research*, 39(1), 104-112.

Taylor, R. (2012). *Analysis of the trends in early entry at GCSE*. Manchester: Assessment and Qualifications Alliance.

Taylor, R. C. (2016). The effects of accountability measures in English secondary schools: early and multiple entry to GCSE Mathematics assessments, *Oxford Review of Education*, 42(6), 629-645, DOI: 10.1080/03054985.2016.1197829.

Taylor, R. C. (2018). *Early and multiple entry to GCSE mathematics and the implications for examination standards* (Unpublished doctoral thesis). Oxford: University of Oxford.

Taylor, R. and Opposs, D. (2018) 'Standard setting in England: A levels'. In Baird, J., Isaacs, T., Opposs, D. and Gray, L. (Eds) *Examination standards: how measures & meanings differ around the world*. London: UCL IOE Press

Theodossin, E. (1986). *The Modular Market. Studies in Further Education*. Further Education Staff College, Blagdon., England. ISBN-0-907659-51-9

Tomlinson, Sir Mike (2004). *14-19 Curriculum and Qualifications Reform. Final Report of the Working Group on 14-19 Reform.* October. Department for Education and Skills Publications. DfE-0976-2004. Available at: http://webarchive.nationalarchives.gov.uk/20050301194752/http://www.dfes.gov.uk/14-19/documents/Final%20Report.pdf

Torrance, H. (2002) *Can education really raise educational standards?* Delivered at the University of Sussex, 11 June 2002. Available at: www.enquirylearning.net/ELU/Issues/Education/HTassess.html

Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (Eds.) *But what does it mean? The use of effect sizes in educational research,* pp. 55–66.

Tymms, P., Merrell, C., & Henderson, B. (1997). The first year at school: A quantitative

investigation of the attainment and progress of pupils. *Educational Research and Evaluation*, 3(2), 101–118. Available at: https://doi.org/10.1080/1380361970030201

Vidal Rodeiro, C. L. & Nádas, R. (2011). *Modularisation: does one size fit all?* Cambridge, Cambridge Assessment.

Vidal Rodeiro, C. L. & Nádas, R. (2012). Effects of modularity, certification session and re-sits on examination performance. *Assessment in Education: Principles, Policy & Practice* 19(4), 411-430.

Vidal Rodeiro, C. L. (2012). An investigation on the impact of GCSE modularisation on A level uptake and performance. *Research Matters,* 2, 21-28. Cambridge, Cambridge Assessment.

Warwick, D. (1987). *The Modular Curriculum*. Oxford, Basil Blackwell.

Warwick, D. (Ed.) (1988). *Teaching and learning through modules*. Oxford: Basil Blackwell.

Watkins, Peter Rodney. (1987). *Modular approaches to the secondary curriculum*. Longman for the School Curriculum Development Committee.

Wheadon, C. (2010). The impact of re-sitting on the fairness of grading. RPA_10_CBW_RP_023. CERP, Guildford.

Wolf, A. (1995). *Competence-Based Assessment.* Buckingham: Open University Press.

Wolf, A. (2002). *Does Education Matter? Myths about education and economic growth*. London, Penguin.

Wolf, A. (2011). *Review of Vocational Education – The Wolf Report*.

YouGov (2019). *Perceptions of A Levels, GCSE and Other Qualifications in England – Wave 17*. An Ofqual commissioned report. April. Ofqual/19/6491/1.

Young, M. (1995). Modularization and Outcomes Approach: Towards a strategy of curriculum of the future. In: J. Burke (Ed.) *Outcomes, Learning and the Curriculum. Implications for NVQ's GNVQ's and Other Qualifications,* pp. 169-181. London, Falmer Press.

Young, M. & Leney. T. (1997). *From A levels to an advanced level curriculum of the future*. Dearing and beyond: 14–19 qualifications, frameworks and systems, 40-56.

Appendix A: Relevant GCSE Policy Changes

Year	Policy description
2000	QCA regulatory arrangements specify: Specifications with staged assessment must normally allocate a weighting of 50% to external assessment through terminal examination(s). The remaining 50% may be allocated to either external or internal assessment, subject to any minimum and/or maximum internal assessment requirement stated in the relevant subject criteria. Where assessment is staged, each assessment unit may be retaken once only. Candidates may, however, retake the whole qualification more than once.
2001	GCSE qualification and subject criteria revised.
2002	GCSEs in vocational subjects introduced (to replace GNVQ Foundation and Intermediate qualifications). Modular in format, with three equally weighted units. Most had 66% coursework. (DfES, 2003) http://webarchive.nationalarchives.gov.uk/20130401151715/http://www.education.gov.uk/publications/eOrderingDownload/DfES0809200MIG1065.pdf
2004	QCA statutory regulations specify that assessment for GCSE should: allow only one retake of assessment units with staged assessments with the better result counting towards the qualification.
2005	*Education and Skills White Paper* sets out plans for Functional Skills qualifications in English, mathematics and ICT, Diploma qualifications and more 'challenging' content to GCSEs.
2005	New science criteria for courses starting in 2006: minimum 25% internal assessment; minimum 25% external assessment; no terminal assessment rule (QCA, 2005, *Science: Changes to the curriculum from 2006 for Key Stage 4*). Different examination boards modularise science in different ways, with different weightings.
2007	GCSE qualification and subject criteria revised. Unitised specifications must: i. allow only one re-sit of an assessment unit with the better result counting towards the qualification ii. allocate a weighting of at least 40% to terminal assessment iii. ensure results for a unit have a shelf-life limited only by the shelf-life of the relevant specification Almost all GCSE *submissions* were modular, although the qualifications criteria at first were neutral as to whether or not they should be. One re-sit allowed; 40% of the assessment had to be terminal; controlled assessment replaces coursework. Proportion of controlled assessment was stipulated for each subject and could only be 25% or 60% (or 0%).
2009	Following critical report of science specifications, changes made to science *examinations* in advance of changes to criteria for GCSEs taught from 2011.
2009	First teaching modular GCSEs (except English, English language, English literature, mathematics, and ICT) (September).
2010	First teaching modular GCSEs in English, English language, English literature, mathematics and ICT (September).

Year	Policy description
2010	*Importance of Teaching* White Paper proposed changes included a return to exams taken at the end of the course, and measures to improve the assessment of spelling, punctuation and grammar.
2010	EBacc measure introduced and first (retrospectively) calculated as part of suite of published performance measures.
2011	First two-year modular GCSEs awarded in all subjects except the English subjects, mathematics and ICT.
2011	Revised GCSE criteria (Ofqual) state that unitised specifications must: • contain a maximum of four assessment units in a single award • allocate a weighting of at least 20 per cent to each assessment unit • allow only one re-sit of an assessment unit with the better result counting towards the qualification • allocate a weighting of at least 40 per cent to terminal assessment • ensure that results for a unit have a 'shelf life' limited only by that of the specification
2012	First two-year modular GCSEs awarded in the English subjects, mathematics and ICT.
2012	Following a review by Professor Alison Wolf, DfE announces reduction in the numbers of non-GCSE qualifications eligible for inclusion in performance measures and that no qualification will count for more than one GCSE in size.
2012	Subject discounting introduced: multiple entries in the same subject but in different types of qualification could no longer be included in performance tables (for 2014), with only one qualification being counted.
2013	Government announced its policy expectations for GCSE, including that they will be linear qualifications and tiering and non-exam assessment to be kept to a minimum in a letter to Ofqual (February).
2013	Government announced that only student's first GCSE entry will count toward school's achievement and attainment (September).
2013	Government announced Progress 8 as a performance table measure from 2016 (October).
2013	Performance tables 2013-2014 only count two non-GCSEs.
2013	Ofqual sets out 9 to 1 grading proposals for consultation (November).
2013	DfE published subject content for new GCSE English language, English literature and mathematics. Ofqual announces assessment requirements for these subjects and outcomes of its consultation on new GCSEs, including that they will be linear qualifications, graded 9 to 1, and principles for use of tiering and for non-exam assessment (November).
2014	DfE publishes subject content for new GCSEs in other EBacc subjects; Ofqual announces associated assessment requirements (April).
2014	Ofqual requires that from summer 2014, regardless of how they were designed, all GCSEs will be taken in a linear fashion – i.e. all examinations must be taken at the end of the course and students must aggregate all units required to complete the qualification at the same time at the end of the course.
2014	Ofqual confirms grading arrangements for subjects to be graded 9 to 1 from 2017 (September).
2014	Performance tables reflect first entry rule for examinations taken from 2013 onward.

Year	Policy description
2015	DfE publishes subject content for new non EBacc GCSE subjects to be taught from 2015. Ofqual announces associated assessment requirements (January and February 2015).
2015	Following its 'Completing reform' consultation Ofqual confirms decisions on those subjects for which exam boards could begin to develop content for first teaching in September 2017 and those existing subjects which will not be reformed as GCSEs (March and May).
2015	Following consultation, Ofqual confirms how practical skills will be assessed in new science GCSEs.
2015/ 2016	DfE publishes subject content for GCSE subjects to be taught from 2016. Ofqual announces associated assessment requirements (November and December 2015 and February and March 2016).
2016	Progress 8 performance measure is introduced for results in qualifications in 2016.
2016	Ofqual confirms grading arrangements for subjects to be graded 9 to 1 from 2018 onwards and revisions to its approach how to grade 9 will be set for all subjects (September).
2017	DfE confirms its ambition that 75% of year 10 pupils in state-funded mainstream schools will start to study GCSEs in the Ebacc combination of subjects by September 2022 and 90% of year 10 pupils studying GCSEs in the EBacc subjects by 2025 (July).
2017	First award of new GCSEs in English language, English literature and mathematics (August).
2018	First award of 19 new GCSEs taught first from 2016 (August).
2019	First award of 25 new GCSEs taught first from 2016 (August).
2020	First award of 5 new GCSEs taught first from 2017 (August). All GCSEs awarded as reformed qualifications.

Appendix B: Unanticipated interplay between qualification design features: GCSE English 2012

- **Changes to the structure of the qualifications**: The newly modularised GCSE English/English language specifications were available for first certification in June 2012. Each specification contained a mixture of written papers (40%), controlled assessment (40%), and speaking and listening assessment (20%), with each module being available prior to the first certification opportunity. Whilst some schools took advantage of the new modular structure and entered their students to modules in the early series (January 2011, June 2011 and January 2012), others entered all of their students in the series that they certificated. For the former schools, students were required to meet the terminal rule and enter 40% of the assessment in the series that they certificated. However, what was not anticipated was the number of students who would enter the written papers early and then use the controlled assessment as the terminal assessment. A one re-sit rule was put in place as well as having a substantial part of the assessment at the end of the course. A compounding factor was that the examination boards developed the unitised qualifications to allow the controlled assessment (coursework) units for English and English language GCSE to be identical, leading teachers to believe that the cut scores, or grade boundaries, would be identical for both.

- **Changes to the number of syllabuses**: Prior to 2010 there were two syllabuses – English, which covered the entire programme of study, and English literature. Beginning in 2010 students in England had three available syllabuses – English, English language and English literature, the first two of which had common coursework. Weaker students tended to be entered for English; stronger students for the combination of English language and English literature.

- **A new programme of study**: The new programme of study was streamlined with less prescription for teachers but also less common material between English language and English literature. It included a new emphasis on English language skills within real-life contexts. Spelling, grammar and presentation skills were highlighted. When content changes, teachers have to modify their lessons and create new ones. In addition they cannot be as certain as when teaching on an older qualification how all of the content will be assessed. Added to this, there was a required weighting of 45% to 55% English and English language attributed to elements of Functional Skills English, emphasising

functionality and application in 'purposeful contexts and scenarios that reflect real-life situations' (Ofqual, 2011, p. 15).

- **Shifting patterns of entry**: Schools could offer either International GCSE or GCSE English for performance table recognition, the outcome of which was 23,000 fewer students from independent and selective schools taking GCSEs in 2012 than in 2011. This meant that the 1.5% reduction in headline statistics for those being awarded a grade C or above was not a like-for-like comparison with the previous year.

- **Increased weighting of controlled assessment (coursework)**: Whereas coursework had counted for 40% (20% speaking and listening and 20% writing) prior to 2010, it now either had to be 25%, which subject specialists believed was too little, or 60%, which might overbalance the internally assessed aspects. The latter was deemed more in keeping with the subject demands, with speaking and listening remaining at 20% of the overall grade. Historically, as well as in 2012, candidates have received high marks for the speaking and listening component. The evidence produced is largely ephemeral, which makes it difficult for examination boards to moderate and adjust the grade boundary marks if necessary.

- **The challenge of setting standards in new qualifications**: The examination boards and Ofqual were faced with the difficult task of setting overall qualification standards for the first time in January 2012. Setting standards is a much more challenging exercise than maintaining them, because of the newness of the qualification and the effects that can have on teacher preparation and implementation and on student response. While using a combination of examiner judgment and statistics – as is done in the maintenance of standards – when initially setting standards, awarders have to make decisions on student performances that may be different from past performances. This is because, as with GCSE English, structure and content may be different and weightings of constituent parts may have changed. The challenge in this case was exacerbated by the fact that the initial standard setting for the early modules was made on very few pre-summer 2012 entries.

- **Using a comparable outcomes framework to maintain standards**; Comparable outcomes means that all things being equal, based on information on prior performance, the proportion of students who obtained certain grades this year should be the same as the proportion who achieved them last year. Setting comparable outcomes is largely a statistical, normative exercise. A comparable outcomes approach keeps

the proportion of grades stable. In the case of GCSE English 2012 the final outcomes were more or less in line with the expected comparable outcomes predictions. However, the grade boundaries, especially for written controlled assessments, were notably different. Had the grade boundaries for the June assessment been the same as for the January ones, there would have been a 20% increase in the proportion of students gaining a grade C and above (Ofqual, 2012), which might have strained the credibility of the qualification, especially given that students tend to do worse, rather than better, in the first year of a qualification.

- **Teacher behaviour**: Teachers naturally wanted as many students as possible to obtain that iconic C grade and unlike GCSE mathematics, English is an open-ended, essay-based subject where the best marking criteria in the world cannot (and should not) exactly describe the entirety of student performance. Because of the modular nature of the new qualifications teachers knew far more about what they believed to be the likely outcomes of their students than in the past. Most students took the externally assessed examination paper and speaking and listening assessments before the summer of 2012, so their marks up until then would have been known prior to the date by which teachers had to submit their marks for the controlled writing assessment. The other 'known' element was the outcomes of the January 2012 standard setting, for which the cut scores for the controlled writing assessment awarded in January were published. English coursework cut scores for grades in the past had been remarkably stable and there was no obvious reason that teachers would believe that things would be different this time around. But there is no guarantee that a certain number of marks will push a student over a grade boundary – cut scores can, do and did change between January and June awarding.

Appendix C: Summative assessment in international examination systems

Jurisdiction	Lower secondary	Format	Upper secondary	Format
Australia: New South Wales	Yearly literacy and numeracy assessments in Grades 7 and 9 (NAPLAN). Essential Secondary Science Assessment in Year 8.	linear	Higher School Certificate exams for each subject worth 50% of students' grades.	linear
Australia: Queensland	Yearly literacy and numeracy assessments in Grades 7 and 9 (NAPLAN).	linear	Externally moderated school-based assessment tasks Grades 10-12. Queensland Core Skills Test Grade 12	continuous assessment linear
Canada: Alberta	Provincial Assessment Tests (PAT), at Grade 9 in English, maths, science, social studies and French. Being replaced by Student Learning Assessments (SLAs) 'readiness' assessments given at the beginning of the school year in Grade 9.	linear	Students enrolled in science 30, biology 30, chemistry 30, physics 30, ELA 30-1 and 30-2, Français 30-1, French language arts 30-1, maths 30-1 and 30-2, social studies 30-1 and 30-2 must take diploma examinations.	linear
Canada: Ontario	Assessment of maths in Grade 9.	linear	Literacy assessment (in order to graduate from high school).	linear
China: Hong Kong	Territory-wide System Assessment (TSA) Secondary 3 in Chinese language, English language and maths.	linear	Hong Kong Diploma of Secondary Education (HKDSE) exams Secondary 6 in Chinese, English, maths and liberal studies plus, typically, two or three other subjects.	linear
China: Shanghai	Graduation exams for lower secondary (*huikao*), ten subject-based examinations. Entrance exams for senior high school (*zhongkao*) in Chinese, maths, English, sciences, social science.	linear	Graduation exams for upper secondary (*huikao*). Entrance exams for university (*gaokao*) in Chinese, maths and foreign language usually English (compulsory) plus sciences or humanities.	linear
Finland			Matriculation exam at end of upper secondary.	linear
France	*Diplôme National du Brevet (DNB)* or *Certificat Formation Générale* consist of final year exam plus marks awarded in class tests over the last two years. Exam consists of three papers in French, maths and history/geography/ civics plus separate IT paper	continuous assessment plus linear final exams	*Baccalauréat Général; Baccalauréat Technologique; Baccalauréat Professional* - all papers externally marked and given once per year.	linear
Germany			*Hauptschule* culminates in *Hauptschulabschluss* after Grade 9 and/or *Realschulabschluss* after Grade 10. *Realschule* culminates in *Mittlere Reife* after Grade 10. *Gymnasium* culminates in *Abitur* after Grade 12.	linear
International Baccalaureate	Middle Years Baccalaureate (MYB) final assessment to determine student achievement levels for each subject group and personal project. Teacher administered. No formal exams.	linear	Six subjects (three at higher level) including extended essay plus creativity, action, service (CAS), and theory of knowledge (TOK)	linear
Ireland (Republic)	Standardised testing in English, maths and science end of second year. Final assessment of subjects at the end of three years set by examination board, marked by schools. Worth 60% of the overall marks. Comprise a single paper or assignment. 40% completed by student during second and third year.	linear	Leaving Certificate exams available at higher and ordinary levels. Students take five or six subjects. Some subjects also have coursework.	linear

Jurisdiction	Lower secondary	Format	Upper secondary	Format
Japan	Diagnostic exams in Japanese and mathematics in 9th Grade entrance exam for upper secondary.	linear	National Achievement test in civics, geography and history, Japanese literature, foreign language, science and mathematics. University admissions exams.	linear
Korea	Some students admitted to senior high school based on lottery. Others take school-administered entrance exam to senior high school.	linear	General High School Diploma or Vocational High School Diploma. Candidates sit seven hours of national exams during one day. College Scholastic Ability Test (CSAT) offered once a year.	linear
Netherlands	In grade 8 schools administer an aptitude test (*Cito* final test primary education) to recommend type of secondary.	linear	Five common subjects plus one specialised subject combination and an independent project. 50% national test, 50% school-based tests.	Combination of school results and nationally set linear assessments
New Zealand	In each subject maximum of three externally assessed standards examined in a three-hour exam.	linear	National Certificate of Educational Achievement (NCEA) qualification. Available at three levels that correspond to the final three years of secondary schooling. At level 3, five subjects.	linear
Scotland	Standard Grades (now National Assessments). Courses divided into elements, usually three units and a final exam. The three units are internally assessed. Units are graded pass/fail and can be re-sat once. Final exam externally assessed and given once a year.	modular with final examination 1x per year	Higher and Advanced Higher (2015) (Higher Still). Courses divided into elements – compulsory core, optional elements, three units and final exam. Three units internally assessed through competency tests drawn from National Assessment Bank plus coursework. Units graded pass/fail and can be re-sat once. Units are qualifications in their own right. Final exam externally assessed and given once a year. It determines the grade.	modular with final examination 1x per year
Singapore	N or O level exams at end of four/five year lower secondary in between six and eight subjects.	linear	H1, H2 and H3 exams at end of upper secondary in around four subjects for academic track students. (H1 first half of H2; graded at AS standard. Can be taken after one year.)	linear
United States: Florida	No Child Left Behind reading and maths in Grade 11, science once in Grades 10 to 12; FCAT science Grade 8; writing Grades 8 & 10;	linear	Florida End of Course Assessment (FLEOC) in algebra I, biology I, geometry, US history, civics. To graduate must pass FCAT Grade 10 reading. ACT, SAT in reading, writing and mathematics; Advanced Placement (AP) subject examinations.	linear SAT and ACT can be taken on multiple occasions
United States: Massachusetts	NCLB reading and mathematics in Grade 11, science once in Grades 10 to 12.	linear	To graduate must have scaled score of at least 240 on Grade 10 ELA and mathematics; scaled score of at least 220 on one of biology, chemistry, physics or technology/engineering. ACT, SAT in reading, writing and mathematics; Advanced Placement (AP) subject examinations	linear SAT and ACT can be taken on multiple occasions

Appendix D: Systematic search of the research literature

Literature inclusion for this study began with keyword searches in five phases. Using the encompassing Oxford library catalogue system and Google Scholar, the first search commenced with these criteria: (*modular* OR *modularity* OR *re-sit* OR *re-sit* OR *end-of-course* OR *terminal*) AND (*examination* OR *assessment*). Without specifying the level of education, this initial search generated literature discussing secondary, vocational as well as higher education with 1,981 results in total. Out of these records, 108 items were identified as relevant and added to our literature database.

The second search phase focused on national assessments in the UK to uncover any relevant literature not using the foregoing keywords. A separate search was conducted using these terms – (*early entry* OR *accelerate* OR *re-sit* AND *GCSE*), resulting in 3,112 results in different combinations of the terms. Out of this pool only 62 items were added to the database because the mass literature about GCSE fell outside the foci of this study. Also, those relevant had already been included in the previous search. The searches in these two phases were not limited to any publication period, i.e. publication time of this body of literature was set as indefinite.

The third phase was conducted through direct inquiry into a few establishments which specialise in assessment research in the country. Exploring the web archives of three organisations, their publications shortlisted by the same aforementioned keywords were imported to our database. These three organisations were Centre for Education Research and Practice (CERP), Cambridge Assessment and The Office of Qualifications and Examinations Regulation (Ofqual). Only literature published between 1978 and 2014 was made available in their online archives, from which 25 publications were filtered out for our literature inclusion. Another valuable source (phase four), *Research evidence relating to proposals for reform of the GCSE,* published by the Oxford University Centre for Educational Assessment (OUCEA) provided an extra seven entries to the literature base.

The exhaustive nature of the list was strengthened by a repeated search in phase five, almost three months after starting phase one. This final search employed four digital databases, namely ERIC, PsycINFO, Scopus and Web of Science. The focus was restricted to 'modular' and 'linear', the two most relevant keywords of this study, formulated as (*modular* OR *linear* AND *exam* OR *assess*) and contributed 20 items to our literature database.

Utilising the expertise of the members in the project team, additional literature was also obtained directly from them. In this way relevant materials not accessible by digital search engines or unavailable in electronic form could be added to our database. Our entire literature base consisted of journal articles, books, book chapters, theses, conference papers and reports. Only research published in English was included in the review. Outside the scope of the review is modularisation of curricula – this review focuses upon modular and linear

assessment. Thus, it includes examinations in the UK and beyond, which might include international tests of English as a second language and assessments that are not composed of standard examination formats (e.g. vocational assessments).

An electronic reference management system, Mendeley, was deployed to catalogue and tag the research literature. Tags were generated mainly in response to the literature found which included: keywords used for the systematic search, examination to which the research referred, level of the education system (e.g. secondary/higher) and some peripheral issues (e.g. gender/teachers' judgment).

Appendix E: Disadvantages and benefits of examination structures

Article	Benefits	Disadvantages
Baird, J. et al. (2009). Students' views of stretch and challenge in A level examinations (p. 41).	- Students could track their progress	
Barham, P. J. (2012). An analysis of the changes in ability and knowledge of students taking A level physics and mathematics over a 35 year period. *Physics Education*, *47*(2), 162–168.		- lack of knowledge retention after taking the modular exam - lower ability of new undergraduate students to perform basic mathematical manipulations and understanding of physics, due to modular A level
Cumming, G. (1990). Modular syllabuses in mathematics - your only possible choice for the 1990s. *Teaching Mathematics and Its Applications*, *9*(3), 111.	- less syllabus content to be 'crammed' than linear examination - students who do not complete a 2-year course can leave with some qualifications - students able to show a record of achievement thus far based on a credit system, and exchange for A or AS grade - retake the examination to improve grade before exchanging for certificate - greater flexibility in teaching and learning of A/AS mathematics; enhance co-teachability of A and AS courses by allowing AS to be taught as a sub-set of A level; flexibility of the order of teaching modules - better use of staffing and resources by schools - bright students who passed GCSE can study some modules earlier - students can change course mid-way, e.g. from pure & applied mathematics to further mathematics - retain the option of linear mode by taking all 4 modules at end of 2nd year	
Department for Education (2012). The evaluation of the impact of changes to A levels and GCSEs (p. 87).	- allows students to achieve along the way, re-sit if necessary - presenting the subject in bite-size chunks, making learning easier; clearer content; fresher memory when taking examination; easier to revise - reduce students' examination stress; learnt a topic before moving on to next	- taking individual modules does not help develop students subject-specific skills, e.g. physical vs human geography as separate modules - model of assessment does not have enough time to develop students' subject skills, e.g. history - does not promote in-depth and longer-lasting learning, particularly on developing subject-specific skills - destroys coherence within subjects, not helping students to see how parts form together - modular may help improve results but not stronger understanding of the subject
Forster, M. (2011). The challenges of ensuring year-on-year comparability when moving from linear to unitised schemes at GCSE. *Research Matters*, (October), 48–51.	- re-sit units to improve grades before getting certificate - improve grade by breaking content into bite-size chunks, but benefits constrained as 40% of assessment has to be taken at end period	- challenges in setting modular grade boundaries, comparability issues - aggregation issues - could have been disadvantaged at top grades on each tier compared to taking linear mode - aggregation negates the benefits of re-sitting (p.49)

Article	Benefits	Disadvantages
		- early candidates are younger than linear counterparts (less mature) - more difficult in judgment decisions about work quality due to unit exams - different purposes and strategies in entering candidates, producing confusing or misleading data info - ensuring parity between units
Goldschmid, B., & Goldschmid, M.L. (1973). Modular instruction in higher education: A review. *Higher Education*, *2*(1), 15–32.	- performance objectives (frequent evaluation facilitates attainment), self-pacing (students choosing their own learning mode) and frequent feedback - allow a variety of instructional activities and topics within a course (more choice) - allow students to identify strengths and weaknesses through remedial modules, repeating, etc., no need to restudy a large amount of content - encourage cooperation and reduce threat of failure; avoid normal curve grading, promoting real achievement by mastery - more choice enhances learning motivation, individualised learning - reduce course redundancy, lessen staff preparation time and encourage module exchange between departments / universities; greater satisfaction to staff (doing the job for what is paid) - staff more able to assist weak learners - more pedagogic freedom to instructors, which encourages scholarly challenges and ideas - students after taking the module can help curriculum design and materials - less subject content, can be tested as a developmental phase - flexible order of taking modules	- cost saved or not remain unclear, but improving learning should be the objective and higher cost can be justified by better learning - more preparation for a module to professors; low incentive in module design as research and publications offer greater rewards - instructors have to prepare for being challenged by more self-pacing nature of module
Graham, T. (2002). AS Mathematics: The results of a survey of schools and colleges. *Teaching Mathematics and Its Applications*, *21*(1), 11–28	- flexible entry patterns - choice of modules - content can be delivered and tested in manageable size - chance of improving grades - feedback gain while working through the course - students can set short-term targets	- retaking exams increases students' examination loads (too many exams have to be taken) - fragmented content, especially with large number of modules - poor modular examination results can also demotivate students
Hart, J. & Howieson, C. (2004). *Unitisation — Benefits and Issues*. *Unitisation - benefits and issues*.	- unitised system beneficial to vocational education sector and companies in improving staff skills or morale, and overall qualification of the workforce - clearly defined learning outcome with popular modules - unitisation makes the qualification system easier to understand, reducing qualifications available - access and progression, higher uptake of qualifications, recognition of prior learning, credit transfer possible - flexibility and responsiveness, making it easier for people to reskill and upskill; quality of learning; quality assurance of qualification, assessment based on standard, not the pace or mode of learning	- simplicity of qualification system not met by simplicity in implementation - assessment of prior learning cumbersome and resource intensive - reliability issues of assessment, leading to complex verification systems - unitisation means burdensome assessment

Article	Benefits	Disadvantages
	- units mobility and recognition, especially moving between employers, admissions to further education or inter-sector transfer	
Hayward, G., & McNicholl, J. (2007). Modular mayhem? A case study of the development of the A level science curriculum in England. *Assessment in Education: Principles, Policy & Practice*, *14*, 335–351.		- problem with developing good and synoptic understanding of science and mathematics, lack of breadth of study - increased examination cost - detrimental impact on quality of teaching and learning - teachers less able to help students to develop a coherent understanding of science subjects, resulting in lack of trust in qualification by higher education sector - accountability and target-driven culture within institutions, employing examination attainment as performance indicator - reinforce instrumental rationality of the learners and carers, linking grades to access to higher education and better career prospect - reinforcing the use of examination grades for commercial values in a market place
Holding, G. (1995). Implementing modular A levels.	- students can demonstrate knowledge more than traditional A level, reducing the uncertainty in the examination system - shorter timescale encourage interim success by maintaining students' motivation; allow students to take a break and resume their studies later (gaps possible) - a step at a time - early and timely feedback to influence students' behaviour, passing one module examination increases motivation of passing more - choice of modules can be outside the subject core - students more motivated, organised, leading to higher retention and better examination results - staff can present content by module instead of whole syllabus - combination of modules, flexibility in cashing in - improved grades help institution to attract students - modular A level links better with other qualifications, e.g. GNVQs - Years one and two can be taught together for some modules	- students may not be working at the required A level standard until much later - modular entry may be made before knowing the results of earlier modules taken - tracking students' module combination is complex - different syllabuses have different numbers of modules and examination periods. Students may be taken away in one subject to sit examination for other subjects - late bloomers less pleased with being pressured early assessment -early module examination may be too soon for some students - modular re-sits can be disruptive and expensive
Marshall, B. (2003). Education: A levels. *Critical Quarterly*, *45*(1-2), 227–236.		- breaking knowledge down into discrete, testable chunks; the whole is more than sum of all parts - difficult to relate exams taken early to others taken at the end of two years

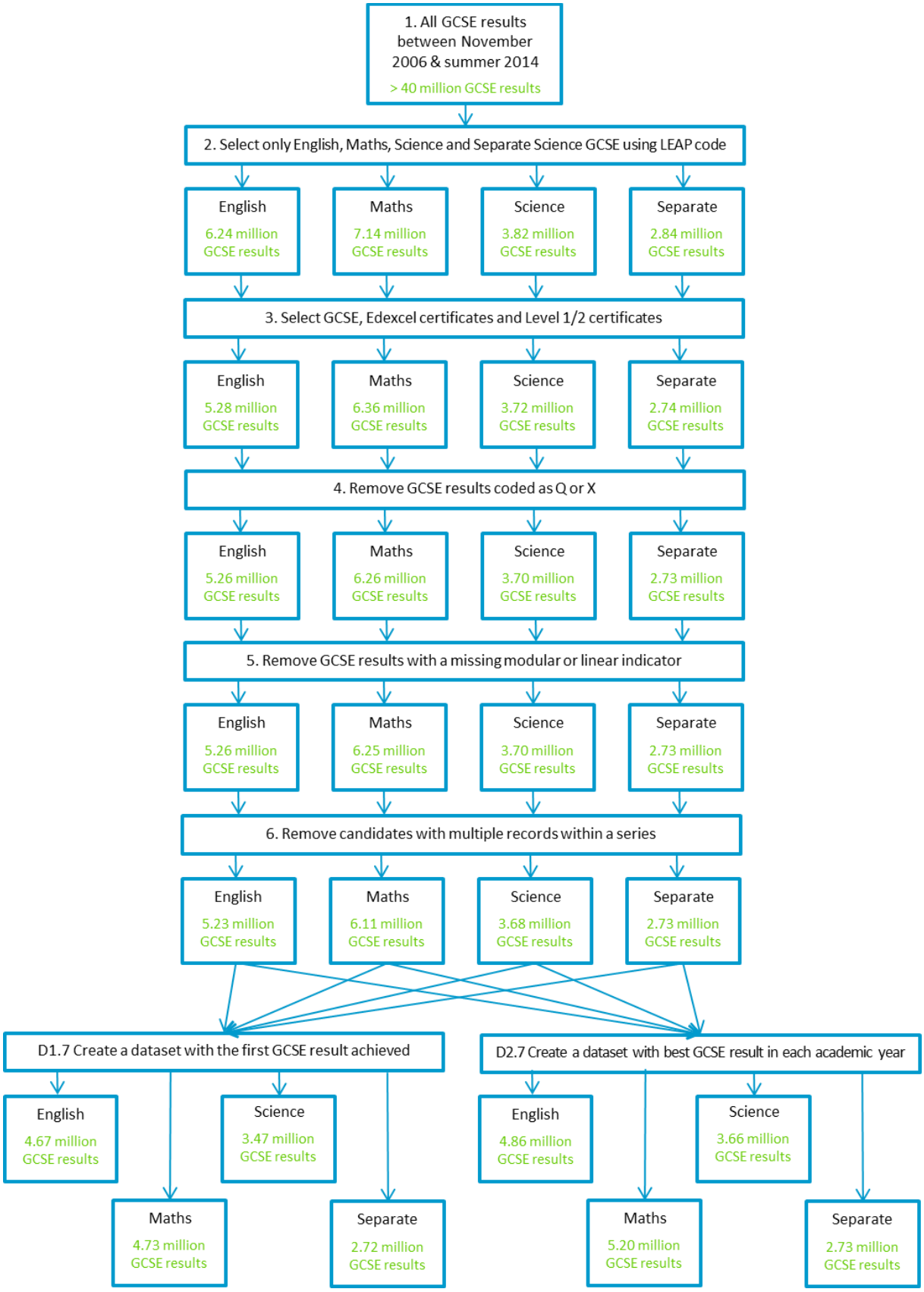
Article	Benefits	Disadvantages
		- syllabuses were constructed around very tightly defined assessment objectives - exams in January mean students go straight from June examination in GCSE to A level within a few months, little time for students to mature, reflect, develop skills outside academic subjects (students placed on a relentless treadmill of exams) - aims of liberal arts education was sacrificed - pupils working hard towards more narrowly defined goals, schools became very utilitarian places - validity of assessment – papers are constructed around what is testable in atomistic form but not what is desirable to test in that subject, e.g. it is difficult to perform to assessment objectives in the mark schemes in English
Noyes, A., & Sealey, P. (2011). Managing learning trajectories: the case of 14–19 mathematics. *Educational Review*, *63*(2), 179–193.	- greater learner self-confidence, improved progress, increase likelihood of future participating in mathematics post-16 - allow students getting higher grades on limited content - keep students motivated - higher grades in GCSE help schools to increase taking-up of A level mathematics	- students less able to see how things connected - increasing grades at GCSE promotes performativity - schools use first A level module as a filter five months into the course, to weed out low attaining students - a concern to fit student interests to maximise engagement, hence, performance - schools tailor the curriculum even more narrowly - modular GCSE did not prepare students for transition to A level as well as linear mode
Pillay, S. (2010). Would the proposed modular assessment for the STPM promote deeper learning or would it erode the quality of Malaysia's pre-university examination? *SEGi Review*, *3*(2), 146–152.	- interactive and timely feedback	- increase teachers' workload - numerous amount of coursework has driven students to surface learning (little time to comprehend and digest the entire curriculum)
Porkess, R. (1997). You can't make a silk purse out of a sow's ear. Why we must say no to the 1997 subject core and its rules. *Teaching Mathematics and Its Applications*, *16*(4), 171–176.	- re-sitting allows more accurate grading, encourages students - allows schools to deal with a wide range of students, from gifted to the less able	- module combinations and weighting are problematic - fragmentation - making modular as like linear as possible is flawed, modular syllabus should be treated in its own right
Vidal Rodeiro, C. & Nádas, R. (2011). The effects of GCSE modularisation: a comparison between modular and linear examinations in secondary education. *Research Matters*, (11), 7–13.	- choice of learning approach, linear or unitised - assessment can be timed to match the point of learning within the course, making students easier to demonstrate their learning - re-sit a unit rather than the entire assessment - timely feedback enables remedy weaknesses - students manage time effectively, revision is more manageable	- modular is associated with low-attaining students - students can be potentially disadvantaged by immaturity or less experience with the subject - fragmentation of learning, lack of coherence - poorly developed overview of subjects, difficult to connect discrete areas of knowledge - assessment dominates throughout the course

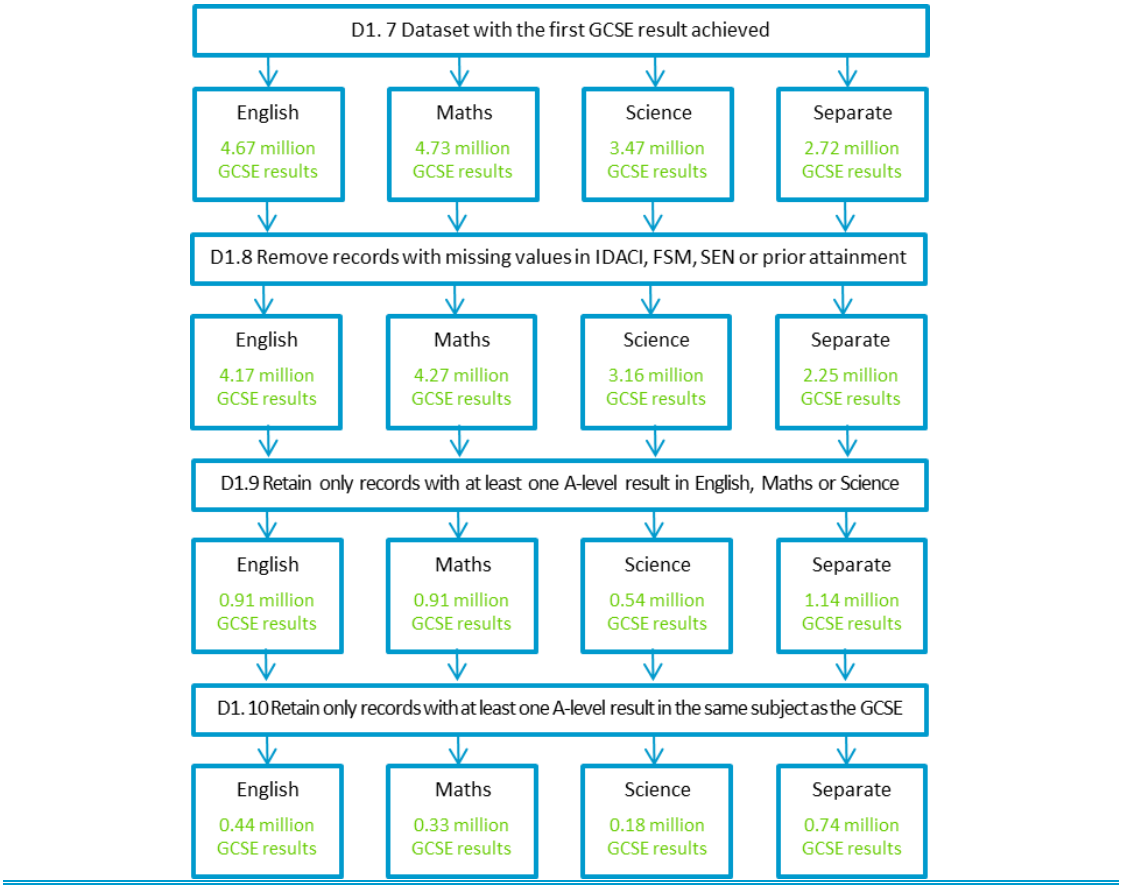
Article	Benefits	Disadvantages
	- assessment load is spread more evenly throughout the two years - assessment more reliable because it is based on more assessed work in total than linear - forge students' sense of ownership - students following modular mode in mathematics perform better than those in linear mode - girls benefit more than boys in modular grades - re-sits release some pressure of 'all-or-nothing' examination - modular exams let students know what to expect in subsequent exams - clarity of focus of teaching requirements to teachers - contributed to teachers' approach to assessment for learning, easier to motivate students for continuous study and revision	- deadlines of units limit teaching of topics teachers see as important - students sit for exams before being ready - encourage a cram-and-discard approach - re-sit unmanaged leads to pressure on school resources and students' workload - less trusted among users of qualifications, higher education, general public, etc. - the choice advantage offered by modular reflect more about the teaching resources or school's preferences rather than students' - modular routes in GCSE English led to lower grades than linear routes - difficult to develop subject maturity - students of modular mathematics experienced longer periods of higher workload than linear students did in 1st half of the year - students were extrinsically rather than intrinsically motivated
Vidal Rodeiro, C. L. V., & Nádas, R. (2010). *Effects of modularisation* (p. 198).	- students in GCSE mathematics following modular route obtained higher grades, can be due to setting targets throughout the course, ongoing feedback, re-taking leading to learning more - modular students are familiar with requirements of exams than those taking the linear mode (provide a sense of 'readiness'), helping them to stay on track with their studies - teachers have better planning around exams, more evenly spread workload throughout the year - fewer 'U' grades in modular than linear route in GCSE English - fit students who are less likely to spend the time later on independent learning and revision, reducing the stress for teachers - grade report provided valuable feedback to students when teachers can discuss common mistakes and make students reflect on their learning. Students learn about strengths and weaknesses, and then work on the weaker spots (even negative feedback helps and can motivate students) - modular syllabuses work differently in different subjects	- students who resat were weaker than those who did not - having re-sits as a choice may have lessened candidates' resolve to do their best at first attempt - modular assessment does not remove the stress and workload of a linear examination - more extrinsic than intrinsic motivation - less control and space for teachers to deliver the content effectively - although no requirements for teachers to enter candidates, teachers may be pressed to do so before finishing the whole course so as to ensure students gaining a pass grade - students in linear route perform better than modular in GCSE English when ability is controlled, and higher probability of obtaining grade A*, and grade B above - negative effects of 'number of units re-sat' on grades obtained – probability of obtaining a given grade decreased if more modules were re-taken - modular students have to revisit material from long-forgotten modules before exams - less flexibility for teachers in timetabling modular exams throughout the year - students in GCSE modular mathematics experienced long periods of higher workload than linear students; the pattern is uneven, suggesting that modular mathematics examination does not in itself alleviate the workload of either linear or modular exams

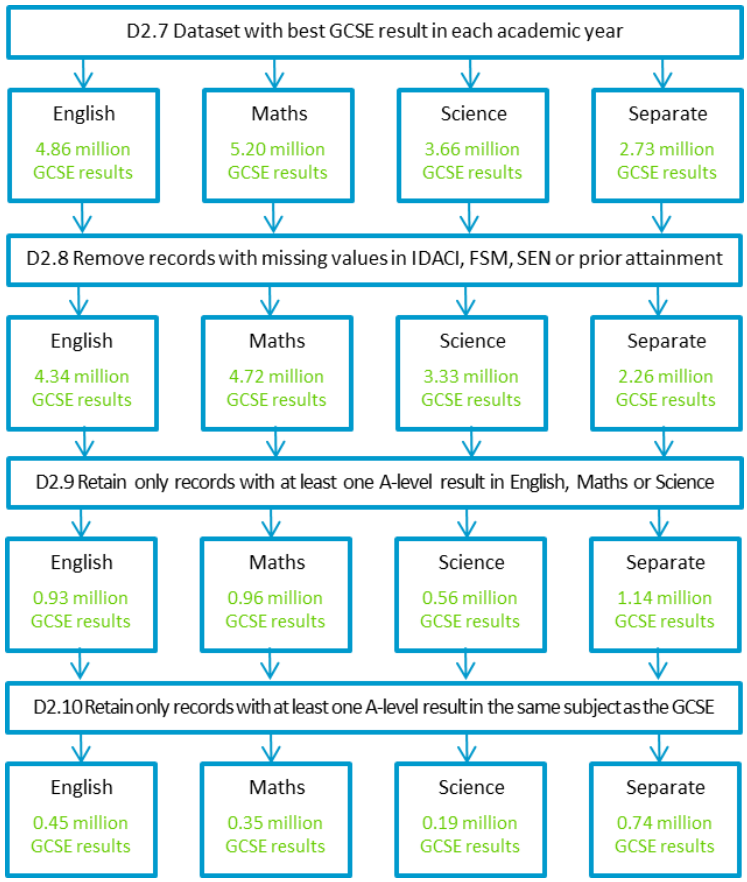
Article	Benefits	Disadvantages
		- modular exams are shorter but more in quantity, making examination process more costly, potentially more disruptive to school routines. - lead to organisational complexity if route design is poor
Scott, E. (2011). *The A levels' re-sit policy and its effect on student learning in three educational institutions in England*. Sussex.	- learning outcomes and assessment criteria are more explicit in smaller units, helping learning and achievement, i.e. enhancing extrinsic motivation - having records of interim performance - provide students a chance to rectify underperformance - students know their interim performance and can try to improve	- over-assessment - fragmented, atomised approach to learning - students fail to grasp the understanding or meaning of the subjects as a whole - teaching was rushed for early exams - the 2-year sixth-form becomes examination-driven - students arriving less prepared for university education even for those with higher scores - students feel that teachers spend most time telling them what they should do in exams instead of teaching the subject, students learn 'learning to forget'
Stringer, N. S. (2012). Setting and maintaining GCSE and GCE grading standards: the case for contextualised cohort-referencing. *Research Papers in Education*, *27*(5), 535–554.		- technical challenge to control subject outcomes compared to linear examination if introduce cohort-referencing – modular candidates carry forward component grades from previous series, making it difficult to control their subject outcomes in certificating series. The system creates less stable outcomes for grade referencing
Thomas, G. (1993). Some Reactions to the Teaching of Science Using a Modular Scheme. *Educational Review*, *45*(3), 213–225.	- setting short-term, clear goals, assessment within credit time span and against explicit criteria, assessment closely articulated with learning – enhance students' motivation - students can choose from a variety of modules - allows a degree of self-assessment by pupils - schools connected with vocational initiative can modularise their timetables - it's a fairer system than having no assessment until the final examination	- teachers concern about time constraints, formal assessment every ten weeks impose burden on teachers - difficult to decide differential weighting of the order or sequence in which modules are taken by different students - establishing standards between modules is difficult faced by certificating bodies - promoting a 'modular curriculum' - if a module is taught by a specialist over 10 weeks, it's difficult for a teacher to get to know a class well, teachers also find it more difficult to maintain discipline - schools running modular courses may result in high staff or student absence, as each module lasts only a few weeks - more difficult to transfer pupils from one stream to another - staff feel that modular system is harder to organise and implement - to teachers, teaching repeated modules is boring, and less long-term contact with students reduce job satisfaction - modular scheme in science became too content-heavy -Teachers feel unable to complete modules in time - a decrease in amount of writing done by students compared to non-modular course, due to more photocopied

Article	Benefits	Disadvantages
		materials provided by teachers in science modular courses - advantages of modular in literature not applied – teachers not involved in preparing for modules, hence little ownership - the system is imposed by the board, not chosen by teachers; illusory benefit of choice to students, as staff choose which modules to present - schools not operating modules in other subjects outside science, the system is unfamiliar to staff and students
Tremain, K. (2011). Findings from focus groups: is controlled assessment working? *CERP Paper*.		- controlled assessment, combined with modularisation, led students to take the assessment less seriously, thinking that they can simply re-take it - students were under pressure from controlled assessment throughout the whole year, worse when combined with modularisation, which increased their workload - students taking the exams in Year 10 are disadvantaged compared to those in Year 11 - lack of opportunities for enrichment in Key Stage 4 timetable

Appendix F: Data sampling flow diagram




D1. 7 Dataset with the first GCSE result achieved
English
4.67 million GCSE results
Maths
4.73 million GCSE results
Science
3.47 million GCSE results
Separate
2.72 million GCSE results
D1.8 Remove records with missing values in IDACI, FSM, SEN or prior attainment
English
4.17 million GCSE results
Maths
4.27 million GCSE results
Science
3.16 million GCSE results
Separate
2.25 million GCSE results
D1.9 Retain only records with at least one A-level result in English, Maths or Science
English
0.91 million GCSE results
Maths
0.91 million GCSE results
Science
0.54 million GCSE results
Separate
1.14 million GCSE results
D1. 10 Retain only records with at least one A-level result in the same subject as the GCSE
English
0.44 million GCSE results
Maths
0.33 million GCSE results
Science
0.18 million GCSE results
Separate
0.74 million GCSE results


D2.7 Dataset with best GCSE result in each academic year
English
4.86 million GCSE results
Maths
5.20 million GCSE results
Science
3.66 million GCSE results
Separate
2.73 million GCSE results
D2.8 Remove records with missing values in IDACI, FSM, SEN or prior attainment
English
4.34 million GCSE results
Maths
4.72 million GCSE results
Science
3.33 million GCSE results
Separate
2.26 million GCSE results
D2.9 Retain only records with at least one A-level result in English, Maths or Science
English
0.93 million GCSE results
Maths
0.96 million GCSE results
Science
0.56 million GCSE results
Separate
1.14 million GCSE results
D2.10 Retain only records with at least one A-level result in the same subject as the GCSE
English
0.45 million GCSE results
Maths
0.35 million GCSE results
Science
0.19 million GCSE results
Separate
0.74 million GCSE results

Appendix G: Patterns of missing data in D2

Table G1 The percentage of missing values for each of the potential covariates in the model

	% Missing			
Covariate	**English**	**Maths**	**Science**	**Separate**
Linear/Modular Indicator	0.0	0.0	0.0	0.0
Academic Year	0.0	0.0	0.0	0.0
Examination Series	0.0	0.0	0.0	0.0
Year-End Age	0.0	0.0	0.0	0.0
Gender	0.0	0.0	0.0	0.0
School Type	0.0	0.0	0.0	0.0
IDACI	0.1	0.1	0.1	0.0
Free School Meals	2.3	1.7	1.8	3.4
SEN (Action Plus/Statement)	5.3	4.1	4.1	7.7
SEN (School Action)	5.3	4.1	4.1	7.7
Prior Attainment (KS2 Average Points Score)	9.9	8.7	8.2	15.9

Table G2 The percentage of data available for analysis in a model of GCSE outcomes controlling for linear/modular assessment route, year-end age, gender, school type, IDACI, a free school meals indicator, SEN indicators and prior attainment

	English				Maths				Science				Separate				
	Linear		Modular		Linear		Modular		Linear		Modular		Linear		Modular		
Year	**W**	**S**	**W**	**S**	**W**	**S**	**W**	**S**	**W**	**S**	**W**	**S**	**W**	**S**	**W**	**S**	**Total**
06/07	95.1	89.9	80.3	68.9	93.6	87.0	91.6	93.4	-	77.9	-	93.4	-	53.3	-	91.6	87.5
07/08	96.2	89.8	80.3	65.8	93.7	87.9	96.0	93.1	92.8	-	89.2	91.0	59.0	93.0	-	75.5	88.2
08/09	95.9	89.1	84.8	66.1	93.5	87.5	94.5	92.5	-	-	91.4	90.6	-	88.9	94.7	78.3	88.1
09/10	95.8	88.8	87.2	65.0	94.3	87.4	94.3	92.5	-	-	90.6	90.7	-	89.8	96.3	81.7	88.4
10/11	95.6	88.6	88.3	61.5	94.8	88.7	95.9	92.0	-	-	92.1	90.6	-	-	92.2	83.3	88.9
11/12	95.7	80.9	93.4	90.9	94.2	90.0	95.0	89.4	-	88.0	93.0	91.9	-	75.5	78.6	85.8	90.3
12/13	92.7	73.0	96.0	91.5	94.9	90.9	95.1	85.9	100	56.2	94.1	92.7	89.6	50.7	91.7	89.5	91.1
13/14	90.3	89.5	95.5	-	94.6	90.6	91.8	-	86.8	91.3	-	-	95.6	87.0	-	-	90.1
Total	**95.7**	**89.2**	**94.4**	**87.2**	**94.5**	**88.8**	**94.7**	**91.9**	**92.7**	**87.3**	**93.4**	**91.6**	**90.3**	**79.6**	**91.5**	**83.8**	**89.1**

Appendix H: The GCSE outcome for each covariate

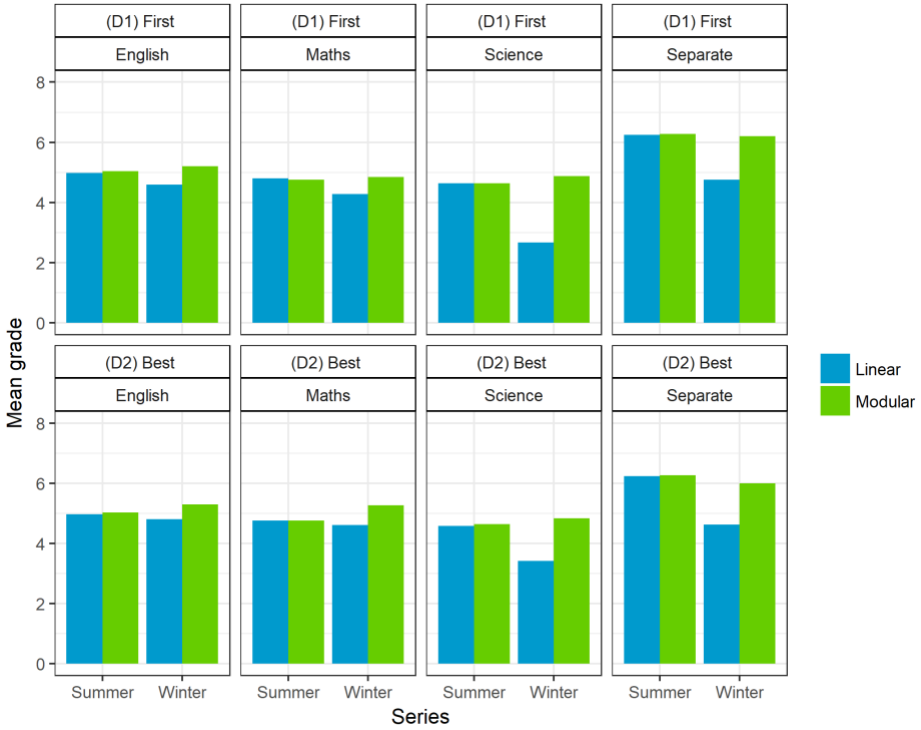


Figure H1 Mean GCSE result by series of entry and assessment route

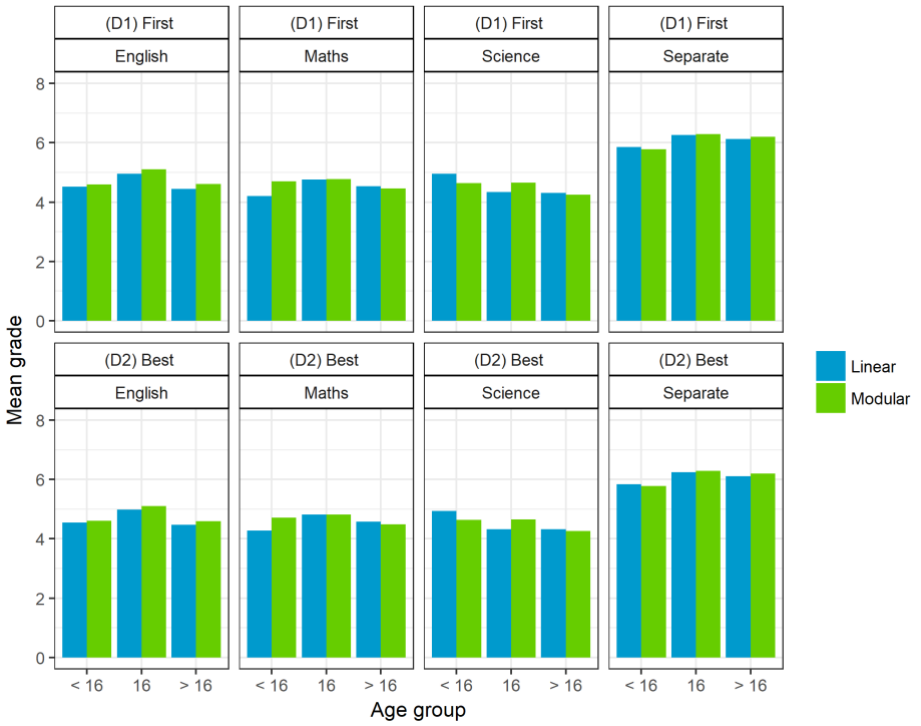


Figure H2 Mean GCSE result by age group and assessment route

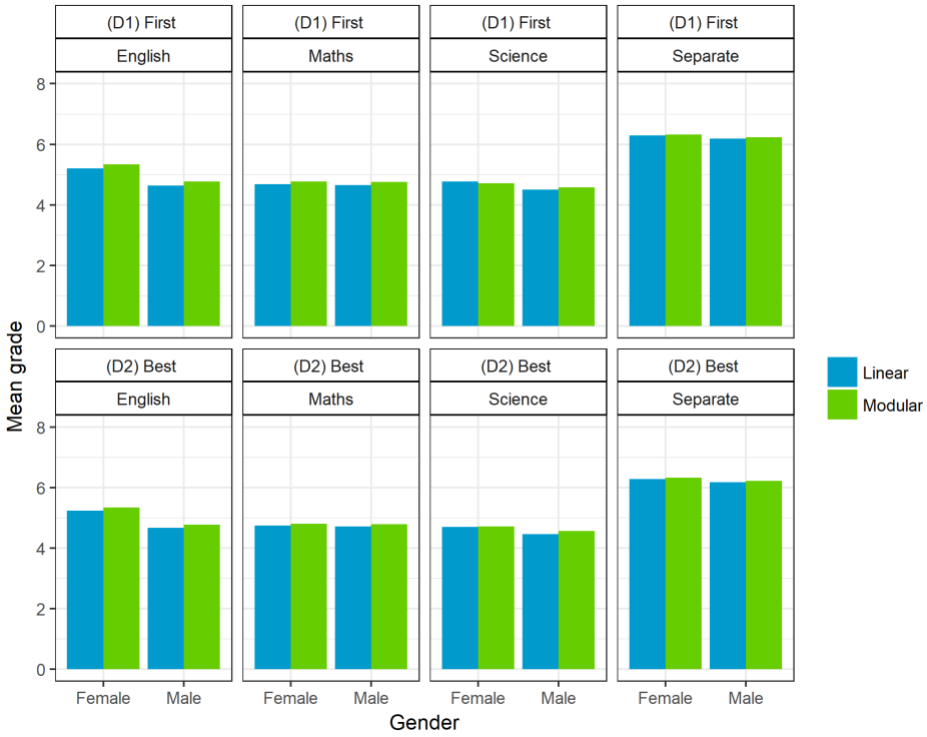


Figure H3 Mean GCSE result by gender and assessment route

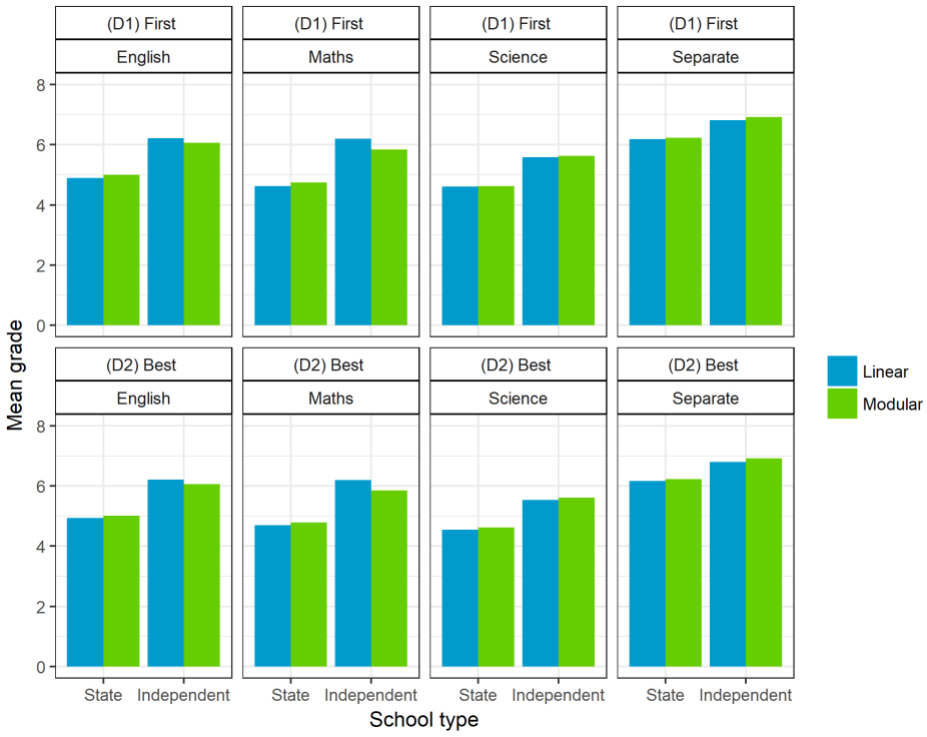


Figure H4 Mean GCSE result by school type and assessment route

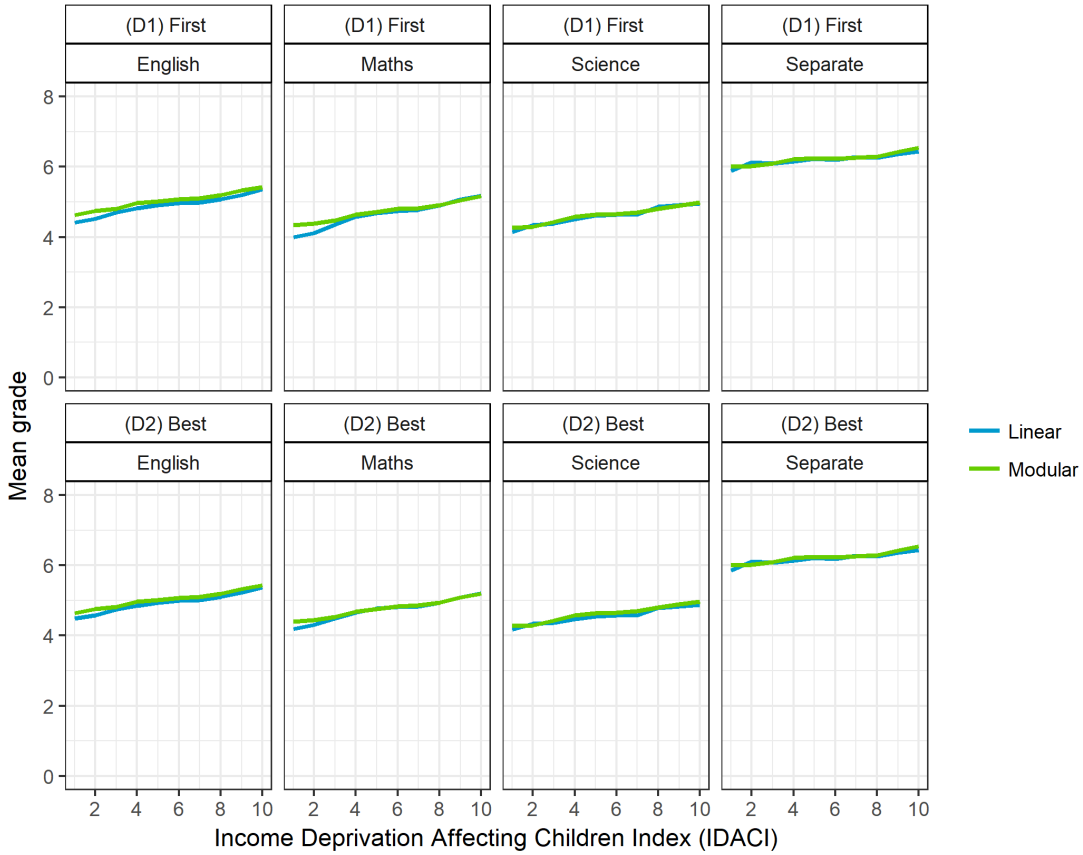


Figure H5 Mean GCSE result by IDACI and assessment route

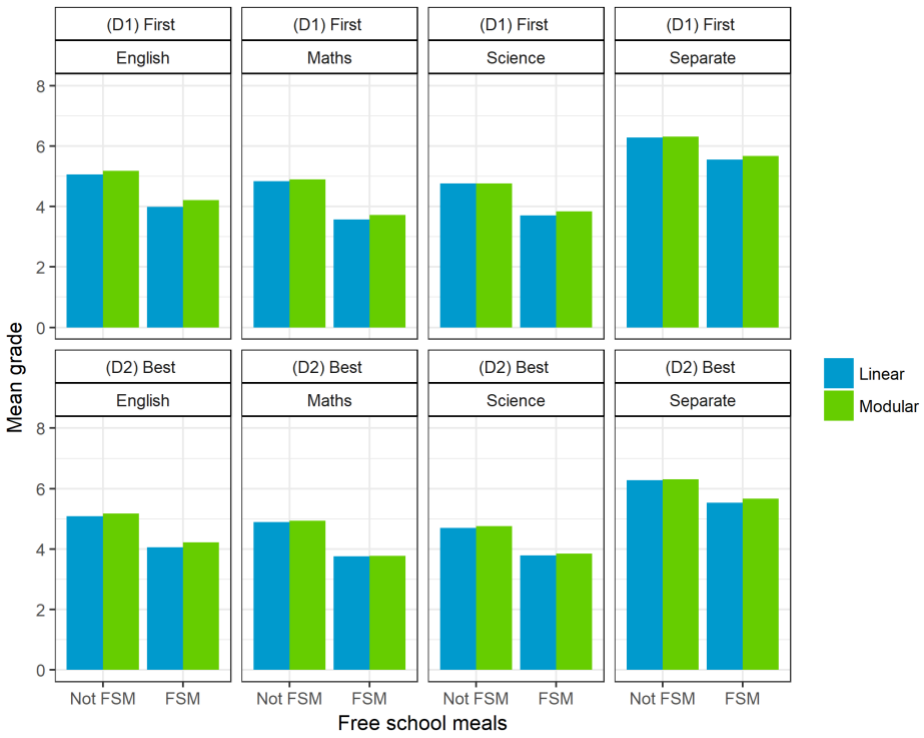


Figure H6 Mean GCSE result by free school meals and assessment route

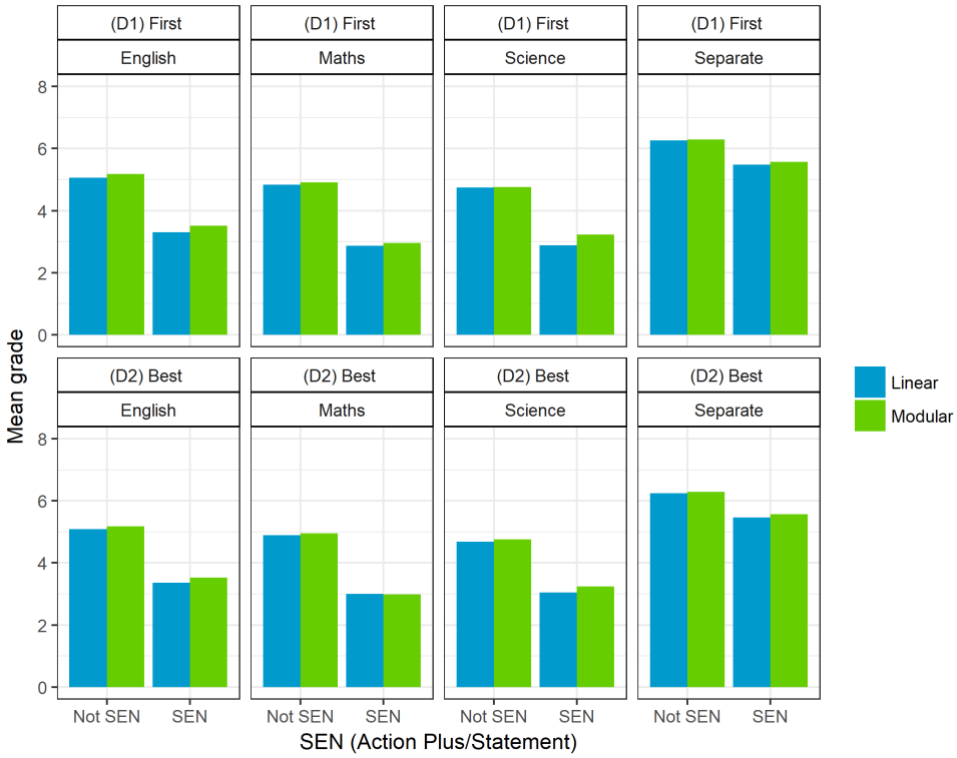


Figure H7 Mean GCSE result by SEN (Action Plus/ Statement) and assessment route

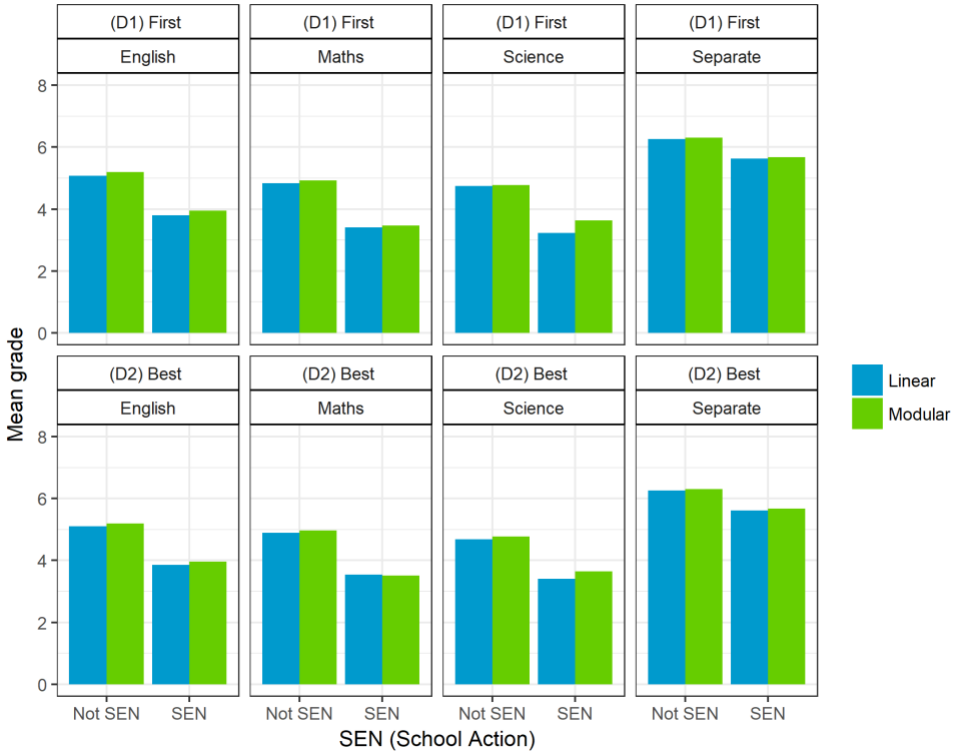


Figure H8 Mean GCSE result by SEN (School Action) and assessment route

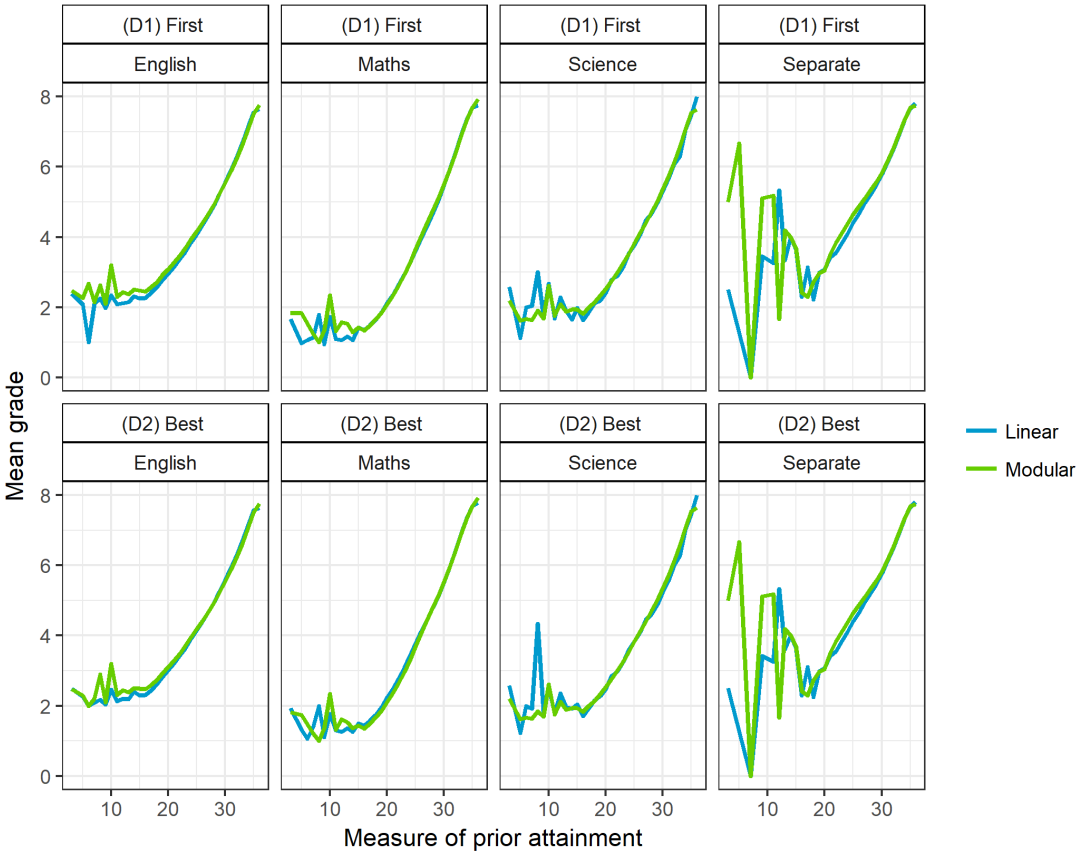


Figure H9 Mean GCSE result by measure of prior attainment and assessment route

Appendix I: Details of the multilevel models fitted to evaluate the effect of assessment route on GCSE outcome

Table I1 Multilevel linear regression analysis of GCSE grade outcomes (D1)

First (D1)	English		Mathematics		Science		Separate	
Fixed Effects	**Estimate (se)**	**Effect Size**	**Estimate (se)**	**Effect Size**	**Estimate (se)**	**Effect Size**	**Estimate (se)**	**Effect Size**
Intercept	5.079 (0.008) †		4.518 (0.008) †		4.543 (0.009) †		6.011 (0.010) †	
Age: <16	-0.270 (0.002) †	-0.276	-0.312 (0.002) †	-0.291	-0.160 (0.002) †	-0.151	-0.212 (0.004) †	-0.222
Age: >16	0.202 (0.010) †	0.207	0.151 (0.011) †	0.141	0.224 (0.014) †	0.212	0.159 (0.017) †	0.166
Series: W	-0.080 (0.002) †	-0.081	-0.113 (0.001) †	-0.105	0.022 (0.005) †	0.021	0.198 (0.015) †	0.207
Gender: Male	-0.467 (0.001) †	-0.476	0.128 (0.001) †	0.120	-0.033 (0.004) †	-0.031	-0.079 (0.003) †	-0.082
IDACI	0.030 (0.002) †	0.170	0.035 (0.002) †	0.183	0.039 (0.002) †	0.207	0.040 (0.003) †	0.226
FSM: Yes	-0.274 (0.002) †	-0.280	-0.236 (0.002) †	-0.220	-0.304 (0.006) †	-0.288	-0.207 (0.007) †	-0.217
SEN (Action Plus/Statement): Yes	-0.671 (0.002) †	-0.685	-0.565 (0.003) †	-0.527	-0.607 (0.009) †	-0.575	-0.253 (0.011) †	-0.265
SEN (School Action): Yes	-0.412 (0.002) †	-0.420	-0.338 (0.002) †	-0.315	-0.543 (0.008) †	-0.514	-0.145 (0.008) †	-0.152
Prior Attainment	1.037 (0.001) †	2.117	1.354 (0.001) †	2.528	1.032 (0.002) †	1.956	0.687 (0.002) †	1.438
Modular	0.031 (0.002) †	0.031	0.057 (0.002) †	0.053	-0.003 (0.003)	-0.003	0.084 (0.003) †	0.088
Prior Attainment ^ 2	0.132 (0.000) †	0.269	0.145 (0.000) †	0.272	0.104 (0.001) †	0.198	0.047 (0.001) †	0.097
Gender: Male * Modular	-0.017 (0.002) †	-0.017	0.002 (0.002) †	0.002	0.057 (0.004) †	0.054	0.020 (0.003) †	0.021
IDACI * Modular	-0.009 (0.000) †	0.000	-0.010 (0.000) †	0.000	-0.006 (0.001) †	0.000	-0.001 (0.001)	0.000
FSM: Yes * Modular	0.045 (0.003) †	0.046	0.021 (0.003) †	0.019	0.042 (0.006) †	0.040	0.032 (0.007) †	0.034
SEN (Action Plus/Statement): Yes * Modular	0.094 (0.004) †	0.096	0.069 (0.004) †	0.065	0.123 (0.009) †	0.116	0.061 (0.012) †	0.064
SEN (School Action): Yes * Modular	0.035 (0.003) †	0.035	0.022 (0.004) †	0.021	0.238 (0.008) †	0.225	0.005 (0.009)	0.005
Prior Attainment * Modular	-0.050 (0.001) †	-0.103	0.076 (0.001) †	0.141	0.076 (0.002) †	0.143	-0.029 (0.002) †	-0.060
Prior Attainment ^ 2 * Modular	-0.009 (0.001) †	-0.018	0.007 (0.001) †	0.014	0.019 (0.001) †	0.036	0.003 (0.001) *	0.007
Random Effects	**Variance**		**Variance**		**Variance**		**Variance**	
School residual	0.227		0.259		0.268		0.277	
Student residual	0.960		1.147		1.114		0.912	
Variance partition coefficient (Null)	35%		34%		31%		35%	
R^2 fixed	0.506		0.580		0.469		0.260	
R^2 fixed & random	0.601		0.657		0.572		0.433	

Table I2 Multilevel linear regression analysis of GCSE grade outcomes (D2)

Best (D2)	English		Mathematics		Science		Separate	
Fixed Effects	**Estimate (se)**	**Effect Size**	**Estimate (se)**	**Effect Size**	**Estimate (se)**	**Effect Size**	**Estimate (se)**	**Effect Size**
Intercept	5.109 (0.007) †		4.615 (0.008) †		4.540 (0.008) †		6.007 (0.010) †	
Age: <16	-0.315 (0.002) †	-0.325	-0.437 (0.002) †	-0.410	-0.154 (0.001) †	-0.146	-0.205 (0.004) †	-0.214
Age: >16	0.199 (0.010) †	0.205	0.157 (0.010) †	0.147	0.223 (0.013) †	0.211	0.153 (0.017) †	0.160
Series: W	0.040 (0.002) †	0.041	0.029 (0.001) †	0.027	-0.001 (0.004)	-0.001	0.073 (0.012) †	0.076
Gender: Male	-0.465 (0.001) †	-0.479	0.122 (0.001) †	0.114	-0.043 (0.003) †	-0.041	-0.080 (0.003) †	-0.084
IDACI	0.025 (0.002) †	0.145	0.019 (0.002) †	0.101	0.036 (0.002) †	0.190	0.039 (0.003) †	0.225
FSM: Yes	-0.266 (0.002) †	-0.274	-0.222 (0.002) †	-0.208	-0.278 (0.005) †	-0.263	-0.211 (0.007) †	-0.220
SEN (Action Plus/Statement): Yes	-0.665 (0.002) †	-0.685	-0.570 (0.002) †	-0.535	-0.562 (0.008) †	-0.532	-0.251 (0.010) †	-0.263
SEN (School Action): Yes	-0.403 (0.002) †	-0.415	-0.338 (0.002) †	-0.317	-0.455 (0.007) †	-0.431	-0.145 (0.008) †	-0.152
Prior Attainment	1.024 (0.001) †	2.110	1.314 (0.001) †	2.467	0.984 (0.002) †	1.865	0.687 (0.002) †	1.438
Modular	-0.001 (0.002)	-0.001	-0.034 (0.002) †	-0.032	0.001 (0.003)	0.001	0.087 (0.003) †	0.091
Prior Attainment ^ 2	0.125 (0.000) †	0.258	0.129 (0.000) †	0.243	0.097 (0.001) †	0.183	0.047 (0.001) †	0.099
Gender: Male * Modular	-0.017 (0.002) †	-0.018	0.007 (0.002) †	0.006	0.067 (0.004) †	0.063	0.022 (0.003) †	0.023
IDACI * Modular	-0.008 (0.000) †	0.000	-0.004 (0.000)	0.000	-0.001 (0.001)	0.000	-0.001 (0.001) †	0.000
FSM: Yes * Modular	0.036 (0.003) †	0.037	0.001 (0.003)	0.001	0.015 (0.005) †	0.015	0.036 (0.007) †	0.038
SEN (Action Plus/Statement): Yes * Modular	0.087 (0.004) †	0.090	0.079 (0.004) †	0.074	0.078 (0.008) †	0.074	0.059 (0.012) †	0.062
SEN (School Action): Yes * Modular	0.028 (0.003) †	0.029	0.022 (0.003) †	0.021	0.151 (0.007) †	0.143	0.004 (0.009)	0.004
Prior Attainment * Modular	-0.046 (0.001) †	-0.095	0.109 (0.001) †	0.204	0.107 (0.002) †	0.203	-0.030 (0.002) †	-0.062
Prior Attainment ^ 2 * Modular	-0.005 (0.001) †	-0.009	0.018 (0.001) †	0.034	0.023 (0.001) †	0.044	0.002 (0.001) †	0.005
Random Effects	**Variance**		**Variance**		**Variance**		**Variance**	
School residual	0.224		0.245		0.265		0.277	
GCSE certificate residual	0.942		1.135		1.114		0.913	
Variance partition coefficient (Null)	35%		33%		31%		35%	
R^2 fixed	0.505		0.573		0.461		0.260	
R^2 fixed & random	0.600		0.649		0.564		0.432	

Table I3 Multilevel logistic regression analysis of GCSE grade outcomes (D1)

First (D1)	English		Mathematics		Science		Separate	
Fixed Effects	**Estimate (se)**	**Effect Size**	**Estimate (se)**	**Effect Size**	**Estimate (se)**	**Effect Size**	**Estimate (se)**	**Effect Size**
Intercept	1.443 (0.015) †		0.430 (0.015) †		0.489 (0.016) †		3.039 (0.023) †	
Age: <16	-0.631 (0.006) †	0.532	-0.649 (0.005) †	0.523	-0.446 (0.004) †	0.640	-0.657 (0.015) †	0.519
Age: >16	0.385 (0.029) †	1.470	0.273 (0.030) †	1.314	0.408 (0.036) †	1.503	0.221 (0.064) †	1.248
Series: W	-0.134 (0.005) †	0.874	-0.114 (0.004) †	0.892	0.093 (0.012) †	1.098	0.665 (0.070) †	1.945
Gender: Male	-0.970 (0.004) †	0.379	0.228 (0.004) †	1.256	-0.051 (0.009) †	0.950	-0.113 (0.013) †	0.893
IDACI	0.057 (0.004) †	1.171	0.069 (0.004) †	1.213	0.065 (0.004) †	1.198	0.059 (0.007) †	1.175
FSM: Yes	-0.515 (0.005) †	0.598	-0.392 (0.005) †	0.676	-0.462 (0.014) †	0.630	-0.384 (0.022) †	0.681
SEN (Action Plus/Statement): Yes	-1.147 (0.007) †	0.318	-0.888 (0.007) †	0.411	-0.915 (0.024) †	0.400	-0.466 (0.034) †	0.628
SEN (School Action): Yes	-0.828 (0.005) †	0.437	-0.557 (0.006) †	0.573	-0.847 (0.020) †	0.429	-0.266 (0.027) †	0.767
Prior Attainment	2.028 (0.003) †	7.599	2.390 (0.003) †	10.914	1.693 (0.007) †	5.434	1.262 (0.010) †	3.532
Modular	0.112 (0.005) †	1.119	0.086 (0.005) †	1.090	-0.078 (0.008) †	0.925	0.232 (0.013) †	1.261
Prior Attainment ^ 2	0.240 (0.002) †	1.272	0.371 (0.002) †	1.449	0.185 (0.005) †	1.203	0.114 (0.003) †	1.121
Gender: Male * Modular	-0.041 (0.006) †	0.960	0.027 (0.006)	1.028	0.102 (0.010) †	1.107	0.052 (0.014) †	1.054
IDACI * Modular	-0.024 (0.001) †	0.935	-0.026 (0.001) †	0.929	-0.011 (0.002) †	0.970	0.004 (0.003)	1.010
FSM: Yes * Modular	0.064 (0.009) †	1.066	0.029 (0.009)	1.029	0.033 (0.015) †	1.034	0.003 (0.024)	1.003
SEN (Action Plus/Statement): Yes * Modular	0.074 (0.012) †	1.077	0.074 (0.013) †	1.077	0.194 (0.025) †	1.214	0.036 (0.037)	1.037
SEN (School Action): Yes * Modular	0.022 (0.009) †	1.022	0.047 (0.009) †	1.048	0.355 (0.021) †	1.427	-0.070 (0.029) †	0.932
Prior Attainment * Modular	-0.030 (0.005) †	0.971	0.153 (0.006) †	1.166	0.244 (0.008) †	1.276	-0.019 (0.011)	0.981
Prior Attainment ^ 2 * Modular	0.019 (0.003) †	1.019	-0.008 (0.004)	0.992	0.045 (0.005) †	1.046	0.000 (0.003)	1.000
Random Effects	**Variance**		**Variance**		**Variance**		**Variance**	
School residual	0.832		0.874		0.733		1.321	
Student residual	3.290		3.290		3.290		3.290	
Variance partition coefficient (Null)	42%		43%		33%		41%	
R^2 fixed	0.536		0.578		0.470		0.239	
R^2 fixed & random	0.629		0.666		0.567		0.457	

Table I4 Multilevel logistic regression analysis of GCSE grade outcomes (D2)

Best (D2)	English		Mathematics		Science		Separate	
Fixed Effects	**Estimate (se)**	**Effect Size**	**Estimate (se)**	**Effect Size**	**Estimate (se)**	**Effect Size**	**Estimate (se)**	**Effect Size**
Intercept	1.523 (0.015) †		0.640 (0.014) †		0.411 (0.015) †		3.011 (0.023) †	
Age: <16	-0.788 (0.006) †	0.455	-0.958 (0.005) †	0.384	-0.404 (0.004) †	0.667	-0.638 (0.015) †	0.528
Age: >16	0.378 (0.028) †	1.460	0.270 (0.028) †	1.310	0.394 (0.034) †	1.483	0.215 (0.064) †	1.240
Series: W	0.246 (0.005) †	1.279	0.353 (0.004) †	1.423	-0.013 (0.009)	0.988	0.115 (0.044) †	1.121
Gender: Male	-0.961 (0.004) †	0.383	0.213 (0.003) †	1.237	-0.071 (0.008) †	0.932	-0.118 (0.013) †	0.889
IDACI	0.045 (0.004) †	1.135	0.044 (0.004) †	1.132	0.058 (0.004) †	1.175	0.060 (0.007) †	1.178
FSM: Yes	-0.505 (0.005) †	0.603	-0.375 (0.005) †	0.687	-0.423 (0.012) †	0.655	-0.385 (0.021) †	0.680
SEN (Action Plus/Statement): Yes	-1.149 (0.007) †	0.317	-0.898 (0.006) †	0.407	-0.800 (0.020) †	0.449	-0.456 (0.034) †	0.634
SEN (School Action): Yes	-0.808 (0.005) †	0.446	-0.550 (0.005) †	0.577	-0.703 (0.016) †	0.495	-0.264 (0.027) †	0.768
Prior Attainment	2.016 (0.003) †	7.508	2.309 (0.003) †	10.060	1.576 (0.006) †	4.833	1.256 (0.010) †	3.511
Modular	0.022 (0.005) †	1.022	-0.096 (0.005) †	0.908	-0.012 (0.007)	0.988	0.247 (0.012) †	1.280
Prior Attainment ^ 2	0.237 (0.002) †	1.268	0.352 (0.002) †	1.422	0.181 (0.004) †	1.199	0.113 (0.003) †	1.120
Gender: Male * Modular	-0.039 (0.006) †	0.962	0.041 (0.006) †	1.042	0.120 (0.008) †	1.128	0.057 (0.014) †	1.059
IDACI * Modular	-0.020 (0.001) †	0.947	-0.014 (0.001) †	0.961	-0.002 (0.002)	0.994	0.002 (0.003)	1.005
FSM: Yes * Modular	0.048 (0.009) †	1.049	0.003 (0.009)	1.003	-0.003 (0.013)	0.997	0.008 (0.024)	1.008
SEN (Action Plus/Statement): Yes * Modular	0.079 (0.012) †	1.082	0.078 (0.012) †	1.081	0.085 (0.021) †	1.089	0.032 (0.037)	1.033
SEN (School Action): Yes * Modular	0.007 (0.009)	1.007	0.037 (0.009)	1.037	0.218 (0.017) †	1.244	-0.070 (0.029) †	0.933
Prior Attainment * Modular	-0.040 (0.005) †	0.961	0.234 (0.005) †	1.263	0.319 (0.007) †	1.376	-0.016 (0.011)	0.984
Prior Attainment ^ 2 * Modular	0.018 (0.003) †	1.018	0.011 (0.004) †	1.011	0.044 (0.004) †	1.045	0.000 (0.003)	1.000
Random Effects	**Variance**		**Variance**		**Variance**		**Variance**	
School residual	0.807		0.786		0.712		1.325	
GCSE certificate residual	3.290		3.290		3.290		3.290	
Variance partition coefficient (Null)	42%		42%		32%		41%	
R^2 fixed	0.533		0.571		0.458		0.237	
R^2 fixed & random	0.625		0.654		0.554		0.456	

Appendix J: Filtering of NPD data

1. Individual GCSE results data were retained for all series from November 2006 to summer 2014 and matched to student level data.
2. GCSE results were selected on the basis of their LEAP code.
   2.1. GCSE English results were selected if they were categorised with a LEAP code of 5010 (English) or 5030 (English language).
   2.2. GCSE mathematics results were selected if they were categorised with a LEAP code of 2010 (mathematics).
   2.3. GCSE science results were selected if they were categorised with a LEAP code of 1310 (science), 1300 (science: single award), 1370 (science: double award), 1390 (science: double award modular), 1010 (biology), 1110 (chemistry) and 1210 (physics). For analysis, the GCSE science results were classified into two groups. The first group included LEAP codes 1300, 1310, 1370 and 1390, which are all combined science qualifications. They are referred to as *science* throughout the report. The second group included LEAP codes 1010, 1110 and 1210. These are separate science qualifications and are therefore referred to as *separate* throughout the report. If the LEAP code was missing, the data were imputed by reference to the qualification accreditation code (QAN) or the examination board subject code. A lookup between LEAP, QAN and subject code was provided by OCR to ensure the best possible comparability with the JCQ results statistics.
3. Data were filtered to include only GCSE full courses[40], Edexcel certificates and Level 1 and 2 certificates. These certificates are more commonly known as IGCSEs®.
4. Data were filtered to remove any records that were awarded a final grade of Q (results incomplete because of an outstanding query) or X (results incomplete because the requirements of the specification have not been met).
5. A flag was attached to each record to indicate whether the GCSE result was obtained from a modular or linear assessment route. Data were filtered to remove any records without a modular or linear flag.
6. Where there were multiple records for a student, in the same series, entered for the same specification, one of the records was randomly selected.

[40] Double awards were also selected for science.

Appendix K: Baseline interview schedule

Interview schedule for heads of department/ head of KS4 in departments

Preamble (by interviewer):
My name is ....., the date is.... and this interview is with the head of maths/ English. As you know we are interested in approaches to linear and modular assessment in schools. We are interested in the way things worked before changes were introduced that meant all students had to take all their modules at the same time (which came into operation in summer 2014), how they work now with the original qualifications, but being taken 'linearly', and the changes you will make for the new specifications from September this year. Can you just tell me how long you've been head of department and what year you started teaching at this school?

1. Can you tell me about how you enter students for GCSE in mathematics/ English? Who do you enter for what, and when do you enter them?

Follow up probes (if not already elicited)

- *Do you make different decisions for different students?*
- *Do targets make a difference?*
- *Is there a policy on re-sitting? Do you see a lot of re-sitting? Has that changed?*

2. What are you planning to do for the new specifications?
3. How has that changed from the past?
4. Have you changed how you teach the material/ subject content for GCSE mathematics/ English?
5. Do you think that different types of student suit different routes?

- *like boys and girls, higher and lower attaining, students from different backgrounds, or students on FSM?*

6. Do you think the change from modular to linear will have an effect on outcomes (results)? For your school? In general?

- *What were your results like last year? Was that different to how they'd been before? Do you think there will be a difference under the new specifications?*

7. What about its having an effect on student motivation?
8. Are there differences in the exam skills that are required?
9. What do you think is behind the policy change from modular to linear?
10. Do you think one of the routes (linear/ modular) is easier?
11a. What do you think the main advantages are of linear exams?
11b. What do you think the main disadvantages are of linear exams?

(If you need it, here is a list of claims that researchers have made about modular and linear exams.)

Finally, is there anything else that you'd like to add?

Interview schedule for head teachers / deputy head in charge of curriculum

Preamble (by interviewer):
My name is ....., the date is.... and this interview is with the headteacher/ deputy head in charge of curriculum/ assistant principal etc. As you know we are interested in approaches to linear and modular assessment in schools. We are interested in the way things worked before changes were introduced that meant all students had to take all their modules at the same time (which came into operation in summer 2014), how they work now with the original qualifications, but being taken 'linearly', and the changes you will make for the new specifications from September this year.

1. Do you have a school entry policy on who gets entered for what and when?
1b. Have you changed your policy in response to the change from modular to linear? In what ways?
2. Do you have a school policy on re-sits?
3. Has the change had any other school-wide effects?
4. Do you think that different types of student suit different routes?
- *like boys and girls, higher and lower attaining, students from different backgrounds, or students on FSM?*

5. Do you think the change from modular to linear will have an effect on outcomes (results)? For your school? In general?
- *What were your results like last year? Was that different to how they'd been before? Do you think there will be a difference under the new specifications?*

6. What about student motivation?
7. Are there differences in the exam skills that are required?
8. What do you think is behind the policy change from modular to linear?
9. Do you think one of the routes (linear/ modular) is easier?

10a. What do you think the main advantages are of linear exams?
10b. What do you think the main disadvantages are of linear exams?

(If you need it, here is a list of claims that researchers have made about modular and linear exams.)

Finally, is there anything else that you'd like to add?

List of claims as prompts, as necessary

a) Modular exams allow students to track their progress.
b) Modular exams encourage students to only learn information short term.
c) Modular exams allow more flexibility in teaching.
d) Modular exams are easier to revise for.
e) Modular exams reduce student stress.
f) Modular exams prevent in-depth and long-term learning.
g) Linear exams give students more time to develop their subject-specific skills.
h) Linear exams promote coherence within a subject.
i) Modular exams increase achievement.
j) They do this at the expense of understanding the subject.
k) Candidates for modular exams are younger and therefore less mature than those entering linear exams.
l) The frequent feedback given by modular exams improves students' performance.
m) Modular exams enable students to set short-term targets which is beneficial.
n) Retakes in the modular system can increase the student's exam load and therefore their stress.
o) Poor modular results demotivate students.
p) Modular exams increase instrumental attitudes in students.
q) Students get higher grades in modular exams because they are being tested on more limited content.
r) Modular GCSEs do not prepare students for A levels as well as linear ones do.
s) In modular qualifications the course is dominated by assessment.
t) In modular qualifications students are encouraged to take exams before they are ready.
u) Modular exams are less trusted by universities and employers.
v) Knowing there's a re-sit opportunity means students don't try as hard.
w) Modular qualifications suit students who aren't good at independent learning and revision.
x) More classroom time is spent on exam preparation for modular exams, which reduces the time available for focusing on the subject.

Appendix L: Linear examinations phase interview schedule

My name is ....., the date is.... and this interview is with the head of maths/ English/ head teacher. As you know we are interested in approaches to linear and modular assessment in schools and how it's affected life in school coming up to this year's GCSE examinations. Can you just tell me how long you've been head (of department) and what year you started teaching at this school?

General

1. What changes have you made in response to the new GCSEs? In relation to: schemes of work; setting or groupings; timetable organisation; matching teachers to classes.
2. Can you identify which changes have been driven by changes to the curriculum and which by changes from modular to linear? (modular schools only)
3. On a curriculum level, have you changed, for example, when you teach which topics, or how long you spend on a topic before moving on? Or how many lessons you have per week? (If so, has that had a knock-on effect on staffing?)
4. Has linearity influenced teaching strategies?
5. Has linearity affected teaching resources?
6. How do you think that these changes have affected your students?
7. Do you think they have affected different groups differently? E.g. boys versus girls, pupil premium or other disadvantaged backgrounds, or lower attaining versus higher attaining? (prompt for motivation as well as outcomes).
8. Do you think that the change from modular to linear will have an effect on results? For your school? In general?
9. Are there differences to the ways that you are teaching exam skills/ preparation?
10. In what ways are linear exams the same as or different to modular?
11. Have your views of linear qualifications changed in the last few years?
12. Is there anything else about the change from modular to linear that you'd like to comment on?

Linear schools only:

13. Do you think the fact that everyone is now doing linear will make a difference for this school? What about for students?

Maths only

14. How are you making the decision to enter students for higher or foundation tier?

15. How do you decide about students on the cusp of the tier?
16. What factors do you take into account?
17. When is that decision made?
18. Have any of these changed with the new GCSE?

English only

19. Has the removal of tiering had any effects on teaching?