Early and multiple entry to GCSE mathematics and the implications for examination standards

Thesis submitted for the degree of
Doctor of Philosophy in Education

Rachel Charlotte Taylor
Kellogg College

Acknowledgements

This thesis has been completed over the past six years of part-time study. During this time I have been employed at the Assessment and Qualifications Alliance (AQA) and Ofqual, and I am indebted to both organisations for giving me the opportunity to undertake this work and for funding me throughout.

I am also extremely grateful to the many individuals that have supported me during this time – in particular to Steve Strand and Jo-Anne Baird for their supervision and expertise; to Michelle Meadows and Ian Stockford for convincing me it was a good idea in the first place; to Ian Stockford and Neil Stringer for their internal supervision while I was at AQA; to the AQA Research Committee and AQA research staff for commenting on drafts of the chapters and the qualitative paper; and to STAG for allowing me access to inter-awarding body data for the first set of quantitative analyses. I would also like to thank Kate Kelly and Victoria Spalding for their assistance in completing the qualitative phase of this research, and the individuals and schools that participated and shared their insight and experiences with me.

Finally, I would like to thank Stuart Cadwallader for putting up with me during various stages of me completing this thesis.

Contents

Acknowledgements....1
Contents....2
List of tables....6
List of figures....7
Glossary of key terms and abbreviations....8
Glossary of organisations....11
Abstract....12
Chapter 1 – The assessment system in England and the maintenance of standards....13
Overview....13
The assessment system in England....14
GCSE Qualifications....14
Features of GCSE qualifications....16
Grades....20
School performance measures and accountability....22
Maintaining GCSE examination standards....24
Summary....30
Early and multiple examination entry....31
Definitions of early and multiple examination entry....31
Rationale....35
Contributions....40
Summary of Chapter 1....41
Chapter 2 – Literature review....43
Overview....43
The history of early and multiple entry....44
Early entry for the most able…....44
The rise in early entry....47
Early entry for a wider range of students....48
Early entry coupled with multiple entry....50
Summary of changes in entry trends....54
Why more early entry and multiple entry?....55
Key policies....56
The motives of schools....61
Summary of possible factors contributing to EMEE....63
What are the possible effects of EMEE?....64
The effects on students....65
Summary of possible effects of EMEE for students....79
The effects for schools....80

The effects for awarding bodies 81
Early and multiple entry in other jurisdictions 82
Summary and perceived limitations of existing literature 84
Research questions 86
Chapter 3 – Research design: using mixed methods 87
Overview 87
Research design and methodology 88
The use of a mixed methods approach 88
Philosophical assumptions 89
The benefits of a mixed methods approach 90
The challenges of a mixed method approach 91
The use of secondary datasets 93
The use of interviews 95
Ethical considerations and reflexivity statement 98
The power relationships in the research 99
The values and position of the researcher 104
Recruitment and incentives 108
Sponsorship of the research 110
Summary of Chapter 3 112
Chapter 4 – Trends in early entry to GCSE 113
Overview 113
Introduction 113
Method 115
Data collection 115
Results 122
The number of early entries 122
The performance of early entry students 129
The prior attainment of early entries 131
Early entries by school type 133
The performance of early entries while accounting for prior attainment 139
Discussion and limitations 143
Summary of Chapter 4 145
Chapter 5 – Schools' entry approaches to GCSE mathematics assessments 147
Overview 147
Introduction 147
Methodology 148
Participants 148
Data collection 151
Analysis 153

Results 154
Drivers 155
Approaches 162
Implications 179
Disquiet amongst teachers 190
Responsibility 193
Discussion 195
Entry approaches 195
Preparing students for assessments and implications 197
Summary of Chapter 5 198
Chapter 6 – Multiple entry and the implications for outcomes 200
Overview 200
Introduction 200
Measurement error in assessment 202
Aims and rationale 210
Limitations of the analyses 211
Methodology 213
Results 216
Single or multiple entrants? 216
What are the characteristics of students following alternative entry routes? 220
Single entrants 229
GCSE mathematics outcomes 233
Discussion 240
Summary of Chapter 6 242
Chapter 7 – Simulating the potential impact on outcomes 244
Overview 244
Introduction 244
Methodology 246
Results 254
Discussion 257
Limitations of simulations 258
Assumptions 259
Summary of Chapter 7 261
Chapter 8 – Examination standards: the implications of early and multiple examination entry 262
Overview 262
What do ‘examination standards’ mean? 262
Defining examination standards 264
Theoretical considerations of examination standards 264

Comparable outcomes and the meaning of examination standards 267
GCSE mathematics entries between 2007 and 2014 270
The rise in early and multiple examination entry 270
The decline in early and multiple entry 272
Implications for examination standards 276
How can standards be maintained when early and multiple entry increase? 276
How can standards be maintained when early and multiple entry declines? 293
Summary of Chapter 8 304
Chapter 9 – Discussion 307
Introduction 307
Review of findings 308
Increases in early entry to GCSE English and mathematics 308
The influence of performance measures on schools' entry approaches 310
Rising early and multiple entry = rising outcomes? 315
Implications for maintaining examination standards 318
Contributions 321
Challenges and limitations 322
Focus of the research 322
Methodological challenges 323
Contextual challenges 324
Further research 326
Concluding comments 328
References 330
Appendix A – overview of key policies 355
Appendix B – development of GCSEs 358
Appendix C – marking and moderation of GCSE assessments 359
Externally assessed examinations 359
Non-examination assessment 359
Appendix D – generating statistical predictions 360
Appendix E – early entry trends for EBacc subjects 362
Appendix F – qualitative paper published in Oxford Review of Education 368
Appendix G – interview schedule 387
Appendix H – transcripts and code book 390
Appendix J – simulation output 396

List of tables

Table 1. Tolerances on matched outcomes compared to predictions 29
Table 2. Similarities and differences between GCSE mathematics and English (2007-2014) 37
Table 3. KS2 scores and 2011 profile of 16-year-old students (n = 601,421) 119
Table 4. GCSE English percentage of entries by age group (n = total entries) 123
Table 5. GCSE mathematics percentage of entries by age group (n = total entries) 123
Table 6. AQA GCSE English A percentage of entries by age group and tier (n = total entries) 128
Table 7. AQA GCSE Mathematics A percentage of entries by age group and tier (n = total entries) 128
Table 8. GCSE English cumulative percentage outcomes for 15-year-olds 130
Table 9. GCSE mathematics cumulative percentage outcomes for 15-year-olds 131
Table 10. Percentage of 15-year-old GCSE students matched to prior KS2 attainment 131
Table 11. GCSE English % of 15-year-old entries by school type (n = total entries) 135
Table 12. GCSE mathematics % of 15-year-old entries by school type (n = total entries) 135
Table 13. Final sample of schools 151
Table 14. Number of students entering on single or multiple occasions 217
Table 15. Cross-tabulation of examination series for double entry students 218
Table 16. The prevalence of multiple entry, mean entry attempts and maximum entry attempts by first grade 223
Table 17. Cumulative percentage outcomes (best grade) by number of entry attempts 224
Table 18. Logistic regression predicting single or multiple entry 229
Table 19. Examination series entered for single entry students 230
Table 20. Cumulative percentage outcomes by examination series for single entrants 232
Table 21. Cumulative percentage outcomes of all vs. best entry attempts 235
Table 22. Cross-tabulation of grades for double entry students 239
Table 23. Cumulative percentage outcomes of both vs. best entry attempts 240
Table 24. Example output from simulations with 5% multiple entry 255
Table 25. Age profile (%) for GCSE mathematics cohort in summer 2013 and 2014 (JCQ, 2014) 294
Table 26. Entries (%) to modular and linear specifications summer 2013 and 2014 295
Table 27. Predictions for the 2013 – 2014 academic year based on results in the 2012 – 2013 academic year (n = 446,384 in 2013 – 2014) 303

List of figures

Figure 1. Example routes through a modular qualification 18
Figure 2. Grades available by tier in tiered GCSE qualifications 21
Figure 3. Awarding committee judgements of script evidence 28
Figure 4. Overview of research design 89
Figure 5. KS2 profile of 15-year-old GCSE English entries between 2007 and 2011 133
Figure 6. KS2 profile of 15-year-old GCSE mathematics entries between 2007 and 2011 133
Figure 7. Mean % of early entries for English and mathematics by school type 135
Figure 8. Mean percentage of early entries for English and mathematics by FSM eligibility 137
Figure 9. Mean % of early entries for English and mathematics by school performance 138
Figure 10. 15- and 16-year-old GCSE English A foundation and higher tier mean mark by KS2 score 141
Figure 11. 15- and 16-year-old GCSE Mathematics A foundation and higher tier mean mark by KS2 score 142
Figure 12. 15- and 16-year-old GCSE Mathematics B foundation and higher tier mean UMS mark by KS2 score 142
Figure 13. Framework for key themes 155
Figure 14. Mean KS2 profile of students by number of entry attempts 226
Figure 15. Mean KS2 profile of single entry students by examination series 233
Figure 16. Mean difference between highest and average grades by number of entry attempts 237
Figure 17. Distribution of true scores on test A from which students are selected 249
Figure 18. Distribution of students' true scores for students entering test A 251
Figure 19. Differences between all grades and students' best grade by proportion of multiple entrants 256
Figure 20. Cumulative % outcomes (best grade) and the proportion of multiple entrants – grade C 257
Figure 21. Outcomes for single and multiple entry students with the same prior attainment 278
Figure 22. Prediction based on a student's final (i.e. highest) grade 299

Glossary of key terms and abbreviations

15-year-olds	The students who reach age 15 during a given academic year. Students aged 15 would be expected to be in Year 10.
16-year-olds	The students who reach age 16 during a given academic year. Students aged 16 would be expected to be in Year 11.
17-year-olds	The students who reach age 17 during a given academic year. Students aged 17 would be expected to be in Year 12.
Academic year	In England the academic year ends on 31st August. For example, the 2012 – 2013 academic year ran from 1st September 2012 to 31st August 2013.
A level	Advanced Levels (also known as General Certificate of Education or GCE). The main qualification taken by students aged 18 in England, Wales and Northern Ireland.
Awarding	The process via which grade boundaries (or cut-scores) are set. Often known as standard setting.
Certificate (verb)	The term is used here to refer to students completing their qualification. The term can also be used to refer to the formal acknowledgement of an individual's result.
Cohort	The population of students in a given academic year. The term may also be used to refer to a specific group of students – for example, a specific age group (e.g. the 16-year-old cohort) or to a group of students sitting a specific qualification (e.g. the GCSE mathematics cohort).
Early entry	Entering an assessment earlier than expected. In this thesis early entry refers to instances where students certificate at GCSE earlier than expected (specifically during Year 10 or mid-way through Year 11).
EMEE	Early and multiple examination entry. The term is used in this thesis to refer to students certificating during Year 10 or midway through Year 11 and on multiple occasions before the end of Year 11.

Examination series	GCSEs are available at set times each academic year, often termed examination series. The largest examination series is in May and June each year.
External assessment	External assessments are typically examinations. They are usually marked by the exam board so are external to the school.
GCSE	General Certificate of Secondary Education. The main qualification taken by students aged 16 in England, Wales and Northern Ireland.
Key grades	The grade boundaries that are set during awarding. At GCSE the key grade boundaries are A*, A, C and F. Grade boundaries that are not key grade boundaries are calculated arithmetically.
Key Stage 2	The four years of schooling in England that ends in Year 6 (aged 11). Students in England sit assessments at the end of Key Stage 2 in English and mathematics.
Key Stage 3	The three years of schooling in England that ends in Year 9 (aged 14). Students in England sat assessments at the end of Key Stage 3 in English, mathematics and science until 2008.
Key Stage 4	The two years of schooling in England that ends in Year 11 (aged 16). Students are generally expected to sit GCSE assessments at the end of Key Stage 4.
Linear assessment	In linear assessment students sit all of the assessments in the same examination series at the end of the course of study.
Modular assessment	In modular assessment the overall assessment is broken down into smaller parts. Students are able to sit (and generally re-sit) each part of the assessment throughout the course of study (see also terminal rule).
Multiple entry	Entering an assessment more than once (see also re-sitting). In this thesis multiple entry refers to instances where students certificate at GCSE on multiple occasions before the end of Year 11.
NPD	National Pupil Database. A dataset held by the Department for Education containing information on all students in England.
Re-sitting	Taking an assessment more than once (see also multiple entry).

Specification	The term is used to refer to different versions of the same qualification in a given subject, e.g. awarding bodies might develop multiple specifications for the same GCSE subject.
Terminal rule	The requirement for students to sit a certain percentage of a modular qualification in the series that they certificate.
Tiered assessment	Assessments that are tiered are aimed at different levels of ability. In GCSEs the higher tier is targeted at grades A* – D (with an allowed grade E), and the foundation tier is targeted at grades C – G.
Year 9	The academic year in which students are expected to reach age 14
Year 10	The academic year in which students are expected to reach age 15.
Year 11	The academic year in which students are expected to reach age 16. Students are generally expected to take GCSE qualifications at the end of Year 11.
Year 12	The academic year in which students are expected to reach age 17.

Glossary of organisations

AQA	Assessment and Qualifications Alliance
CCEA	Council for the Curriculum, Examinations and Assessment
DfE	Department for Education, a ministerial department responsible for children's services and education
Edexcel	Pearson
JCQ	Joint Council for Qualifications, an umbrella body providing a single voice for examination administration
OCR	Oxford, Cambridge and RSA
Ofqual	The Office of Qualifications and Examinations Regulation, a non-ministerial department that regulates qualifications, examinations and assessments in England
Ofsted	The Office for Standards in Education, Children's Services and Skills, a non-ministerial department that is responsible for inspecting a range of educational institutions
WJEC	Welsh Joint Education Committee
QW	Qualifications Wales, the exams regulator in Wales

Abstract

GCSEs are high stakes assessments in England generally taken by students aged 16. The results are put to several uses: they allow students access to further opportunities and are used to judge the performance of schools. While students are expected to take GCSEs once aged 16, this thesis explores an alternative approach – early and multiple examination entry (EMEE). This approach rose in prominence between 2007 and 2013 for GCSE English and mathematics, before changes to accountability measures in September 2013 meant that single entry once again became the norm.

A multiphase mixed methods design was used focusing on EMEE in GCSE mathematics. The first (quantitative) phase showed that early entry increased from 2007 to 2011, generally due to weaker students and students from lower performing schools increasingly entering early. The second (qualitative) phase suggested that the rise in EMEE was due to the pressure of performance measures driving behaviour in schools, encouraging EMEE and associated teaching practices. The third (quantitative) phase suggested that EMEE was likely to have implications for examination outcomes, potentially leading to increased outcomes (i.e. higher grades) at the cohort level.

The findings are considered in terms of the implications for examination standards, drawing on the theoretical literature relating to how examination standards are defined. While the comparable outcomes approach in England implies a causal definition of standards, the rise and subsequent decline of EMEE challenges this definition, raising questions about how standards can be maintained in the presence of EMEE.

Although the prevalence of EMEE has now declined in England, many of the issues identified here are likely to transcend EMEE and have wider implications. Thus, this thesis contributes to the literature relating to the effects of accountability measures on schools' behaviour and how examination standards are defined, in addition to adding to the existing, albeit limited literature, focusing on EMEE.

Chapter 1 – The assessment system in England and the maintenance of standards

Overview

High stakes tests, assessments that have important consequences for the test-takers or other key stakeholders, are common to many assessment systems (Stobart and Eggen, 2012). This includes the assessment system in England, which is centred upon qualifications taken at specific points in a student's education – the General Certificate of Secondary Education (GCSE) typically taken at age 16, and the A level typically taken at age 18. The results from these qualifications are put to a number of uses: they allow access to further progression opportunities for individual students, and are used in accountability measures to judge the performance of schools.

While students are expected to take GCSEs and A levels at specific points in their education, the purpose of this thesis was to study one alternative approach – early and multiple examination entry (referred to as EMEE throughout this thesis) to GCSE assessments. This approach rose in prominence between 2007 and 2013, before changes to accountability measures in September 2013 meant that single examination entry at age 16 once again became the norm.

The thesis focuses on the rise and fall of EMEE and consider trends in EMEE, the drivers of EMEE, and the implications of EMEE for maintaining examination standards. This introductory chapter therefore provides the necessary context to allow readers to engage with the issues that are discussed in this thesis. This includes an overview of the assessment system in England, focusing on GCSE qualifications, and a discussion of how examination standards are maintained.

The work presented in this thesis was completed between 2011 and 2017, as the story of EMEE was unfolding. While it has always been possible to enter assessments early, such cases were generally rare and it is between 2007 and 2013, as this work progressed, that

this approach became increasingly popular. This required the focus of this thesis to evolve and adapt as education policy and entry trends to GCSE assessments changed. As such, the thesis is presented as real-world research; a narrative over time, rather than a comprehensive study that was conducted at a particular time point once the whole story was known. This means that the analyses presented in the early chapters of the thesis have not been updated to reflect additional data from later years. Indeed, it is unlikely that much of the work presented in this thesis would have been conducted at all, had it not been done at the time that EMEE was increasing. The thesis therefore has been written to capture and preserve the story of EMEE as it unfolded.

The assessment system in England

The following description of the assessment system focuses on the period that the work presented in this thesis relates to – between 2007 and 2014 – and therefore does not refer to the most recent round of GCSE and A level qualification reform (for qualifications first taught in some subjects from 2015). The description focuses on the GCSE qualification and predominantly relates to the assessment system in England. Although GCSE qualifications are also available in Wales and Northern Ireland, the assessment systems in the three countries have become increasingly distinctive in recent years (see Office of Qualifications and Examinations Regulation [Ofqual], 2017a). Furthermore, the wider context within which the qualifications operate differs between the three countries, particularly in terms of school accountability systems, which is particularly pertinent to the work presented here. A timeline summarising the key policy changes referred to in this thesis is provided in Appendix A.

GCSE Qualifications

The assessment system in England centres upon qualifications that are typically taken at a certain point in a student's education – the General Certificate of Secondary Education (GCSE) typically taken at age 16 (at the end of Year 11), and Advanced levels (A levels)

typically taken at age 18 (at the end of year 13),[1] both of which are high-stakes assessments. GCSEs were first introduced in 1988 to replace the former qualifications that were available (O levels and CSEs), and were intended to be sat by the majority of the 16-year-old cohort (Daugherty, 1995).

The GCSE has traditionally been considered a school leaving qualification, although since 2013 all young people aged 17 – 18 must be in some form of education, training or employment, with the majority choosing to remain in education (Department for Education [DfE], 2015a). While this has raised questions around the purpose of the GCSE more recently, the majority of students aged 16 continue to take GCSEs in a number of subjects, with the results being put to various uses. At the student level, GCSE results allow students access to further study (e.g. A level courses) and training or employment opportunities, while at the institution level they are used to judge the performance of schools.

GCSE qualifications are available in over 50 subjects (Ofqual, 2009) and a typical student takes between eight and ten GCSEs. All students study some core subjects (English, mathematics and science) then take a combination of other subjects, chosen by the school or the student (depending on availability at individual institutions). GCSE qualifications are generally studied over a two-year period[2] between ages 14 and 16, though more recently there is some evidence of schools starting GCSEs earlier than this (Office for Standards in Education, Children's Services and Skills [Ofsted], 2017).

---

1 Note that other qualifications are available at each level, but the GCSE and A level are taken by the majority of the cohort.

2 Note that some GCSE subjects are available as a short course qualification (e.g. religious studies and history). Short course GCSEs are half the size of a GCSE and are generally taken after one year of study.

GCSE qualifications are offered by awarding bodies (also known as exam boards), which develop[3] and deliver the qualifications. There are four main awarding bodies offering GCSE qualifications in England: the Assessment and Qualifications Alliance (AQA), Edexcel (Pearson), Oxford, Cambridge and RSA (OCR) and the Welsh Joint Education Committee (WJEC).[4] These exam boards are members of the Joint Council for Qualifications (JCQ), which is responsible for providing a single voice for examination administration.

For large entry GCSE subjects (e.g. English, mathematics and science) each awarding body typically offers one (or two) GCSE qualifications in that subject. Smaller entry subjects might only be offered by one or two awarding bodies. The awarding bodies are regulated by the Office of Qualifications and Examinations Regulation (Ofqual), whose statutory objectives include ensuring that examination standards are secured (Ofqual, 2015a).

Features of GCSE qualifications

Examinations and non-examination assessment

The structure of GCSE qualifications differ depending on the subject, but they typically contain a number of assessments (known as components or units) that are either externally assessed examinations or non-examination assessment (traditionally known as 'coursework' or, more recently, 'controlled assessment').[5] A typical GCSE contains between two and four assessments and the balance between examinations and non-examination assessment varies by subject. Externally assessed examinations are typically 1 – 2 hours long and contain a number of multiple choice, short response, or longer essay

[3] See Appendix B for a description of how qualifications are developed and the various public bodies involved.
[4] Prior to 2012 the Council for Curriculum Examinations and Assessment (CCEA), the exam board in Northern Ireland, also offered qualifications in England.
[5] See Appendix C for a description of how each type of assessment is marked.

style questions, depending on the subject being assessed. Non-examination assessment typically requires students to complete a task over a longer time period under controlled conditions.

GCSE assessment structures

GCSE qualifications have typically adopted a linear or modular structure. In linear qualifications, all assessment is taken at the end of the course and students wishing to re-sit a qualification must re-sit all of the assessment.[6] In modular qualifications, students are able to sit (and re-sit) parts of the assessment throughout the course in smaller 'chunks'. Typically though, there is a requirement for students to sit a certain percentage of the assessment in the series that they certificate (or complete the qualification), generally known as a 'terminal rule'. As an example, for GCSEs taught from 2009 students were required to sit at least 40% of the overall assessment in the series that they certificated (Qualifications and Curriculum Authority [QCA], 2008). Despite this, by their design, modular qualifications allow students to sit (and re-sit) parts of the assessment prior to the series in which they certificate – in the case of GCSEs, this can be up to 18 months beforehand (e.g. students certificating in a June series aged 16 might sit modules in the November of the previous academic year aged 15). This means that there are multiple routes through a modular qualification, and the number of routes increases as the number of examination series increase (see section on the availability of assessments). Figure 1 provides some example routes through a modular qualification for students certificating in summer 2014 (note that there are more possible routes than this).

[6] It is usually possible to carry forward the marks from non-examination assessment from a previous examination series when re-sitting.

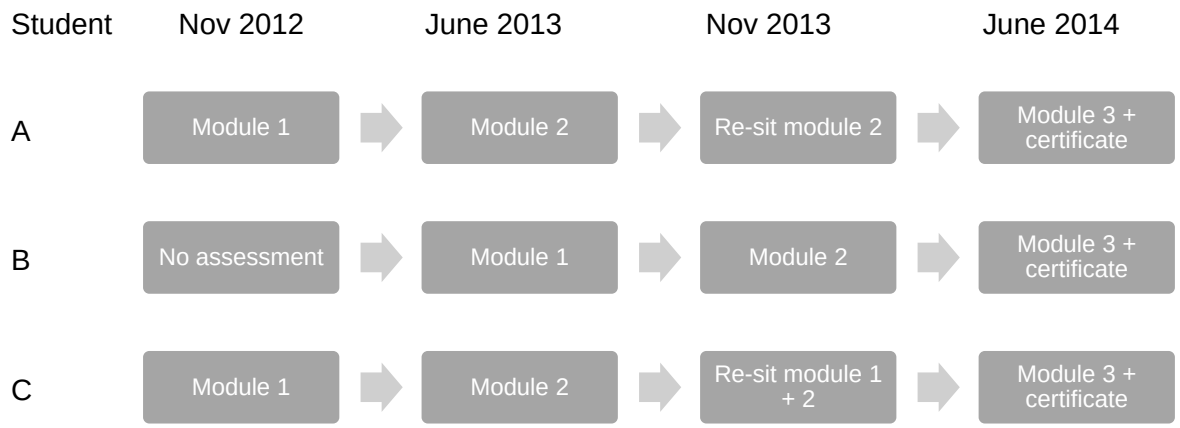


Figure 1. *Example routes through a modular qualification*

The structure of GCSE qualifications has been back and forth in recent years, including during the period that this thesis relates to (a review of the impact of linear and modular assessments at GCSE was underway at the time of writing; see also Vidal Rodeiro and Nádas, 2012). Prior to 2008 the majority of GCSEs adopted a linear structure, requiring students to sit all of the assessment at the end of the course. When new qualifications were developed for first teaching in 2009 (2010 for English, mathematics, and ICT), exam boards chose to develop mainly modular GCSEs. This was allowed, although not required, by the Government at the time. The existence of modular GCSEs was relatively short-lived, however, due to concerns about the amount of assessment and the opportunities to re-sit individual modules (DfE, 2010). As such, all GCSEs returned to a linear format from summer 2014[7] following direction from the Government (DfE, 2010), and have adopted this approach since.

While for most GCSE subjects the assessment has either been linear or modular, mathematics has been relatively unique in that both assessment formats were available concurrently prior to 2014. Thus, schools could chose the approach that they considered

[7] Note that the qualifications were not redeveloped at this point, but all of the modules had to be sat in the same series that students certificated in.

most appropriate for their students, and could even switch between the two throughout the two year course (i.e. by entering students to the early modules of a modular qualification but not certificating in this qualification, instead choosing to enter the full linear qualification; see Vidal Rodeiro, 2014). From 2014, both the linear and (formerly) modular GCSE mathematics qualifications remained available, but all of the assessments had to be sat in the series that students certificated in. This meant that students could not re-sit individual modules of the modular qualification.

Tiering

Some GCSE subjects have traditionally been offered at multiple levels (or tiers), aimed at learners of different abilities. Tiered qualifications generally comprise a higher tier and a foundation tier.[8] Tiering is generally used for subjects that differentiate by task, rather than by outcome. For example, in GCSE mathematics, writing an assessment that includes the range of tasks required to test both the lower and higher achieving learners is considered difficult (see Ofqual, 2013a). As such, the higher tier contains tasks that target the higher ability students, and the foundation tier contains tasks that target the lower ability students. Given this, the grades that students can achieve on each tier differ – the higher tier targets the higher grades, and the foundation tier targets the lower grades. There is also generally a small overlap in the middle so that some grades are available on both tiers (see section on grades).

Availability of assessments

The main examination series for GCSE qualifications, when the majority of students sit the assessments, is in May and June each year, with results issued in August. For English and mathematics there is also a November examination series. This was typically considered as a re-sit opportunity for students who sat the assessment the previous

[8] Note that prior to 2008 GCSE mathematics was available at three tiers – foundation, intermediate and higher.

summer, but prior to 2014 students were able to enter the November assessments as a first attempt (since 2014 November has been a re-sit only opportunity).

Further assessment opportunities have also been available for GCSE mathematics each year. Prior to the end of the 2012 – 2013 academic year, GCSE mathematics was also available in the January and March examination series, although not all awarding bodies offered every qualification in every examination series.[9] Nonetheless, for the majority of the time that EMEE was prominent, it was possible to sit GCSE mathematics in up to four examination series each academic year – this equates to an individual student being able to sit up to 12 assessments for a single GCSE subject in a given academic year.

Grades

GCSEs are graded qualifications. The method for awarding grades differs slightly depending on the structure of the qualification (i.e. whether they are linear or modular). Essentially though, grade boundaries are set each examination series at key grade boundaries, and students receive an overall qualification grade based on their performance when all of the assessments are aggregated together. Where students re-sit a GCSE qualification they receive the highest grade that they achieve as their final grade.

The process of determining grades is known as awarding. The grade boundaries are set each examination series after the papers have been sat by students and marked by examiners, and are able to move each year. This ensures that any differences in the demand of the question papers is accounted for.

[9] The availability of modules in modular qualifications has typically followed a similar timetable, with modules and certification available in March, June and November (January was typically a linear only opportunity).

At the time this thesis was written, GCSE qualifications were graded A* – G[10] (and A levels were graded A* – E),[11] with an unclassified (U) grade given to students not meeting the lowest grade threshold. For tiered qualifications, the higher tier was targeted at grades A* – D (with an allowed grade E for students falling a small number of marks below the grade D boundary), and the foundation tier was targeted at grades C – G (see Figure 2).

Higher	A*	A	B	C	D	(E)			U
Foundation				C	D	E	F	G	U

Figure 2. *Grades available by tier in tiered GCSE qualifications*

While all GCSE grades (A* – G) are considered a 'pass' in the qualification, a grade C has traditionally been considered a 'good' pass (DfE, 2010). This threshold has therefore become increasingly important and is often used as a gateway for allowing students access to employment opportunities. For example, many employers require individuals to have at least a grade C in GCSE English and mathematics.

GCSE grades are also used to allow students access to A level courses and higher education. Many schools and colleges require students to achieve a minimum number of 'pass' grades at GCSE to progress to A levels, and some also require students to achieve a particular grade in a subject at GCSE to progress to an A level in that subject. For example, to study A level mathematics many institutions require students to have achieved at least a grade A or B at GCSE. GCSE results are also used by some

[10] While A levels have retained their grading structure in the most recent round of reform (for some subjects first taught in 2015), GCSEs have adopted a new grading structure of 9-1, with 9 being the highest grade.

[11] The A* grade was introduced in 1994 for GCSE and 2010 for A level to allow better differentiation between the highest achieving students.

universities when making entry decisions to their degree courses, though sometimes less formally. However, this can still raise the stakes of the qualifications for the students.

School performance measures and accountability

The grades that students achieve at GCSE are important at the individual level, but they are also used at the school level to judge the performance of schools in what are commonly known as school performance tables.[12] First introduced in 1992 by the then Conservative Government, school performance tables aimed to hold schools accountable to key stakeholders and provide parents with information about each schools' performance (West, 2010; West, Mattei and Roberts, 2011; West and Pennell, 2000).

Traditionally, the key performance indicator has been the percentage of students in each school achieving five GCSEs (or equivalent) at grades A* – C, although there have been changes to the measure over time (see Acquah, 2012, for an overview).[13] One of the key changes occurred in 2007 when GCSE English and mathematics became mandatory requirements in the five A* – C measure (Department for Education and Skills [DfES], 2005a). While English and mathematics have featured as 'core' subjects in the National Curriculum since the Education Reform Act in 1988 (Department of Education and Science [DES], 1988), their mandatory inclusion in performance tables raised their significance further. These changes sought to ensure that schools were not rewarded for entering students to alternative qualifications that contributed a high tariff towards the five A* – C measure (e.g. see Perryman, Ball, Maguire and Braun, 2011; Wilson, Croxson and Atkinson, 2006), and meant that students would only meet the required threshold if they achieved a grade C or above in both GCSE English and mathematics. This resulted in considerable shifts in the outcomes for some schools (Baker, 2007). Other changes have

[12] Also referred to as performance measures throughout this thesis.
[13] Note that from 2016 the key performance measure changed from the five A* – C threshold measure to progress 8 and attainment 8 – a measure of each student's progress in eight GCSE subjects based on their prior attainment at KS2 (DfE, 2017a).

included restricting the qualifications that can be included in the measure as ‘equivalent’ to GCSE, perhaps most notably following the Wolf review (Wolf, 2011).

Another key change associated with the five A* – C measure occurred in September 2013 when it was announced that, henceforth, only a student’s first examination entry would count in school performance measures (note that students would still receive their highest grade as their final grade though; DfE, 2013a). Prior to this, the highest grade that a student achieved was used. This change had significant consequences for schools’ entry approaches to GCSE qualifications, in particular for EMEE.

In addition to allowing schools to be ranked, the percentage of students achieving five GCSEs at grades A* – C (including English and mathematics from 2007) has been used to hold schools accountable, by identifying schools that are considered to be under-performing via the use of ‘floor standards’. Floor standards specify the minimum percentage of students in a school that are expected to achieve the five A* – C performance standard. Since their inception, these targets have become gradually more stringent (Acquah, 2012), and prior to the changes to accountability measures from summer 2016, schools were considered to fall below the floor if fewer than 40% of their students achieved five GCSEs (including English and mathematics) at grades A* – C. Schools falling below the ‘floor’ could be subject to increased monitoring, inspections and ultimately, closure.

Although the five A* – C threshold has generally been the key performance measure used to judge the performance of schools, other measures have been present as well, perhaps most notably the English Baccalaureate (EBacc) announced in the Schools White Paper in 2010 (DfE, 2010). The EBacc aimed to reward performance in five GCSE subjects that primarily aligned with the Russell Group facilitating subjects at A level: English, mathematics, science, a modern or ancient language, and a humanity such as history or geography (note that the measure has changed over time though; see Armitage and Lau, 2017, for details). While the EBacc has never been a key performance measure,

Government aspirations are for an increasing number of students to take this combination of subjects (90% by 2025; DfE, 2015b, 2017b). The EBacc is therefore of relevance to this thesis since it is another policy that has likely influenced schools' curriculum offering and entry approaches to assessments over the past few years (see Chapter 2 for further discussion).

In addition to the accountability and performance measures described above, schools are also held accountable for the quality of their teaching and students' learning by Ofsted, the Office for Standards in Education, Children's Services and Skills. Founded following the Education Act in 1992, Ofsted is a non-ministerial Government department that is responsible for inspecting, and giving ratings to, a range of educational institutions. While the inspections are typically focused on teaching and learning, at times there has been a greater focus on attainment in examinations (see Perryman, Maguire, Braun and Ball, 2017). Again, this is of relevance to this thesis since it was discussed by some of the respondents in the qualitative phase of this research.

The use of performance measures to judge schools has been, and continues to be, controversial. There is a large body of literature documenting the potential distorting effect that such measures can have on school behaviour (e.g. see Acquah, 2012, for an overview, and Chapter 2 for a discussion). This is particularly the case when performance measures have focused on a threshold measure, such as the requirement for students to achieve five GCSEs at grades A* – C. This focus is considered to have prompted the use of a number of approaches within schools to improve results (see Chapter 2).

Maintaining GCSE examination standards

The structure of the assessment system in England and the uses to which results are put has implications for the methods via which examination standards are maintained (see Newton, 2007, for a discussion). In particular, there is an impetus to ensure comparability of examination results – both over time and between specifications in a single examination

series (see Chapter 8). This is because students achieving the same qualification in different years or from different providers might be competing against one another for education or employment opportunities. The issue of comparability and the meaning of examination standards has long been debated in England and the theoretical literature relating to this is discussed in Chapter 8. Here we focus on how comparability is sought through the process of setting grade boundaries and awarding grades.

The approach to maintaining standards for GCSE (and A level) qualifications has traditionally used a combination of statistical and judgemental evidence, an approach that has been termed weak-criterion referencing (Baird, Cresswell and Newton, 2000) or more recently, attainment referencing (Newton, 2011). Since 2010, examination standards have been maintained using the comparable outcomes approach introduced by Ofqual (2015b).[14] This approach is rooted in earlier work by Cresswell (2003) and subscribes to the 'similar cohort adage' (Newton, 2011), the premise being that if the population of students is similar from one year to the next, then the outcomes should also be similar. As the name suggests, comparable outcomes prioritises comparable outcomes over comparable performance, a feature that protects students sitting assessments in the first year that a new qualification is available (see Cresswell, 2003) when performance is likely to dip – an effect known as the saw-tooth effect (Cuff, Meadows and Black, 2017; Linn, 2000; Ofqual, 2015b).

In practice, comparable outcomes is operationalised through the use of statistical predictions (Ofqual, 2015b). The statistical predictions are prior attainment based predictions that model the relationship between prior attainment and outcomes in a reference year (i.e. a previous examination series), then apply this relationship to the cohort of students sitting the same qualification in the current year (see Appendix D for an

[14] Note that awarding bodies have always used statistical evidence to maintain standards, though prior to 2010 the evidence used was not necessarily consistent across awarding bodies.

explanation of the methodology for generating predictions). The predictions therefore provide an indication of the expected grade distribution for the current cohort of students (at the overall cohort level), given their prior attainment. As such, if the prior attainment of students in the current year is similar to the prior attainment of students in the reference year, then the expected distribution of grades (i.e. the prediction) in the current year will be similar to the outcomes in the reference year. Each awarding body offering a qualification in the same subject must apply the same relationship between prior attainment and outcomes to their cohort of students (see Ofqual, 2017b). This aims to ensure comparability between awarding bodies and is operationalised via the use of national prediction matrices for each subject (see Appendix D).

Prior attainment based predictions are generated for a specific cohort of students. This is usually the age group that is expected to sit each qualification, and therefore the largest group of students – 16-year-olds at GCSE (and 18-year-olds at A level). Since the predictions are based on prior attainment they only include those students that are matched to their prior attainment – often termed 'matched' students. When predicting GCSE outcomes students are matched to their prior attainment on tests taken at the end of Key Stage 2 (KS2),[15] and when predicting A level outcomes students are matched to their mean GCSE score. Grade boundaries are set on the basis of the matched students then applied to all students sitting each qualification.

The use of prior attainment based predictions relies on a number of assumptions – in particular, that the same type of students are entering the assessments each year and

[15] Students generally take KS2 tests at the end of Year 6 (aged 11). Prior to 2010 all students took assessments in English, mathematics and science, but more recently students have only taken assessments in English and mathematics. This has had implications for the measure of prior attainment that has been used in the statistical predictions, although the specific details are not discussed here. It is also worth noting that prior to 2008 Key Stage 3 (KS3) assessments were taken by students aged 14 (see Chapter 2). These tests provided the measure of prior attainment that was used in the statistical predictions until 2010 (see Eason, 2010, for a discussion of the use of KS2 results as the measure of prior attainment).

therefore the value added relationship should be the same.[16] For this reason, the reference series (that predictions are based on) is typically the same series in a preceding year, based on the assumption that the same type of students will certificate in the same series each year and that the relationship between prior attainment and outcomes will therefore remain stable. As such, the outcomes for a June examination series would typically be predicted from a preceding June series, while the outcomes for a November series would typically be predicted from a preceding November series, and so forth. This is due to the likely differences in cohorts between the two series – 16-year-old students in a June series are at the end of KS4, while 16-year-old students in a November series are entering early with less studying time and maturity. Thus, the relationship between prior attainment at KS2 and GCSE outcomes is likely to differ for these different types of students.

The prior attainment based predictions drive the setting of grade boundaries but there is also a role for qualitative judgement. As such, senior examiners review students' work around the grade boundaries suggested by the statistical evidence in comparison with archive scripts. These archive scripts are from a previous examination series and represent the required standard of work on each key grade boundary. Examiners are asked to determine whether the work they are reviewing from the current examination series is of the same standard or not as the archive. Scripts are reviewed in this manner in a range around the grade boundary suggested by the statistical evidence, and examiners' views are recorded on what is known as a 'tick chart' (Figure 3). The tick chart is then used in conjunction with the statistical evidence to recommend grade boundaries.

[16] It is for this reason that independent and selective schools are excluded from GCSE predictions, since their value added has been shown to be different (see Eason, 2010).

Mark	Chair of Examiners	Chief Examiner	Principal Examiner A	Principal Examiner B	Principal Examiner C	Principal Moderator
54	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓
53	✓✓✓	✓✓✓	✓✓?	✓?✓	✓?✓	✓✓✓
52	✗?✗	✓✗✗	✓✗✓	✓✗✓	✓✗✗	✓✗✗
51	✓✗✗	✓✗✗	✗✗?	✗✗✗	✗✗✗	✗✗✗
50	✗✗	✗✗	✗✗	✗✗	✗✗	✗✗

Figure 3*. Awarding committee judgements of script evidence*

Grade boundaries are set in this manner for key grade boundaries, then the remaining grade boundaries are calculated arithmetically. The key grade boundaries at GCSE are A*, A, C and F.[17] The grade B boundary, for example, is therefore calculated as the mark that is halfway between the grade A and C boundaries (rounded down if necessary). Grade boundaries are set on each assessment that make up a qualification, then students' results on each assessment are aggregated (though the method for this depends on the structure of the qualification). In linear qualifications, the grade boundaries are generally added together to give an overall subject level grade boundary (though the marks from individual assessments are sometimes scaled), and any student achieving the required number of marks across all of the assessments achieves that particular grade. In modular qualifications, students' marks are standardised using what is known as a uniform mark scale (UMS).[18] This allows students sitting the assessments in different examination series (where the grade boundaries may differ due to the demand of the papers) to be treated equally, i.e. a grade A is always worth 80% of the UMS marks, regardless of how many raw marks are required to achieve a grade A.

[17] A* is a key grade boundary but it is set arithmetically initially, not using judgemental evidence.
[18] See AQA (2016) for further details about the uniform mark scale.

Once exam boards have set the grade boundaries and these have been signed off internally, they report their outcomes to Ofqual, the exams regulator. Ofqual monitors the outcomes from each awarding body compared to the statistical predictions, using a set of tolerances at particular grades (Ofqual, 2017c).[19] The tolerances depend on the size of the matched cohort (i.e. the number of students included in the statistical predictions), since the greater the number of matched students then the more reliable the predictions will be and the smaller the tolerances. The tolerances that are used are shown in Table 1. Where the outcomes for an individual awarding body exceed tolerance they must provide an explanation to Ofqual, who decide whether to accept the outcomes or ask for further evidence. Ultimately, Ofqual is able to direct awarding bodies to make changes to grade boundaries if the evidence provided is not considered sufficient to support the award deviating further from the statistical predictions than the tolerance allows.

Table 1. *Tolerances on matched outcomes compared to predictions*

Number of matched students	Reporting tolerance
500 or less	None
501–1,000	3%
1,001–3,000	2%
3,001 or more	1%

The comparable outcomes approach has been in place since 2011 for GCSE, yet exam boards have always used a combination of statistical and judgemental evidence to guide the setting of grade boundaries. What has changed over time though is how this evidence is balanced (Newton, 2011). Research highlighting potential biases in examiner judgement and the tendency for examiners to give students the 'benefit of the doubt' (e.g. see Baird, 2000; Baird, 2007; Baird and Scharaschkin, 2002; Scharaschkin and Baird,

[19] The reporting tolerances apply to grades A and C at GCSE and grade A at A level. In addition, for subjects with a sufficient number of students there is a 2% tolerance at A*.

2000; SCAA/OFSTED, 1996, cited in Tymms and Fitz-Gibbon, 2001) has led to greater emphasis on the statistical evidence (Baird and Gray, 2016). Thus, within the comparable outcomes approach the statistical predictions drive the process (Ofqual, 2015b), and it is against these predictions that outcomes are evaluated to consider whether standards have been maintained (see Ofqual, 2017c, and Chapter 8 for further discussion).

Grade boundaries for the summer examination series are generally set during July, with results issued to students in mid-August. Both GCSE and A level results are subject to intense public and media scrutiny each year, with national outcomes reported extensively in the media (Warmington and Murphy, 2004; 2007). Small rises in outcomes typically lead to assertions that standards are falling and that the examinations are getting easier (e.g. see Murphy, 2004). This was particularly the case prior to the introduction of comparable outcomes, when GCSE results had generally risen year-on-year (Ofqual, 2015b).

Since the introduction of comparable outcomes in 2011, GCSE outcomes have generally remained stable each summer. Concerns around falling standards have consequently been replaced by assertions that comparable outcomes caps results and does not allow genuine improvements in student performance to be recognised. Thus, it is argued that, although schools are under intense pressure to improve results (and resources are channelled towards this), outcomes are not able to increase in response. While the debate around the possible effects of comparable outcome will no doubt continue (e.g. see Benton, 2016), further discussion of this particular issue is beyond the scope of this thesis.

Summary

In summary, there were a number of features of GCSEs between 2007 and 2014 that are relevant to this thesis: GCSE qualifications were expected to be taken once at age 16; assessments were available multiple times throughout the academic year; multiple qualifications were available in some subjects; there were limited restrictions on re-sitting

the qualifications; accountability measures were focused on a threshold performance measure (i.e. grade C); and English and mathematics were mandatory requirements in performance measures. It is within this context that EMEE was able to prosper.

Early and multiple examination entry

The description of the assessment system outlined above provides an overview of the typical route through the assessment system at the time that the work in this thesis was conducted. While traditionally the majority of students followed this route, between 2007 and 2013 other routes to entering GCSE qualifications became more prominent. One of the ways in which this occurred was via EMEE. This approach rose in prominence between 2007 and 2013 for a wide range of students, including entire cohorts of students in some schools, before effectively ceasing following changes to accountability measures announced by the Government in September 2013 (DfE, 2013a) that meant that only a student's first entry attempt would count in school performance measures.[20]

The practice of EMEE had significant implications for students, schools, and the wider assessment system between 2007 and 2013, and it is the use of EMEE during this period that is the focus of this thesis. Before considering the rationale and specific focus for this research, it is important to consider what is meant by the key terms – early entry, multiple entry and EMEE.

Definitions of early and multiple examination entry

The term 'early entry' has been commonly used in the English assessment context, particularly between 2007 and 2014, yet there are no specific definitions within the literature of what constitutes an early examination entry. Despite this, given the structure

[20] Note that many students entered GCSE English early in November 2013 shortly after this announcement was made. However, the following summer the numbers declined significantly.

of the assessment system and the expectation that qualifications will be sat at certain ages (and in certain school years), it follows that, in its broadest sense, early entry can be defined as instances where students certificate (or complete) a particular qualification before intended.[21] Specifically focusing on GCSEs, this would mean certificating before the end of Year 11 (students are generally aged 16 by the time they complete Year 11).

Within this definition some distinction between the varying degrees of early entry is also required. Students could certificate during (but before the end of) Year 11 in some subjects (e.g. certification in GCSE mathematics was typically available in a January, March and November examination series between 2007 and 2013), or students could certificate before entering Year 11 (most commonly in Years 9 and 10). Furthermore, students might certificate even earlier than this. Each August there are reports of very young students taking GCSEs in the media (Roberts and Bingham, 2010; see also Gill, 2010). However, since these students are likely to represent unusual scenarios with specific motives, they are not the focus of this thesis. Instead, the thesis will focus on early entry when students certificate in Year 10 or mid-way through Year 11, as the evidence suggests that entering prior to this (i.e. in Year 9 and before) is relatively rare (DfE, 2011). As such, the term early entry is used to refer to instances of students entering their GCSE qualification during Year 10 or mid-way through (but not at the end of) Year 11, regardless of how many times they sat the qualification in total and whether they also entered at the end of Year 11.

This brings us to the second key term in this thesis, ‘multiple entry’. Again, there is no specific definition of multiple entry in the literature. However, since students are expected to certificate once aged 16 (i.e. at the end of Year 11) it follows that multiple entry can be

[21] Note that in modular assessments students can (and often do) sit assessments before the series in which they certificate. This thesis is concerned with instances of early certification though, and not early entry to individual modules.

defined as instances where students certificate on more than one occasion.[22] Similar to early entry, there are varying degrees of multiple entry though: students might enter once at the end of Year 11 then re-sit in post-16 education (as many students not achieving a grade C in their first attempt do), or they might enter on two or more occasions before completing Year 11. Furthermore, students might enter on multiple occasions in different examination series, or might enter twice in the same examination series.[23] While there are therefore many routes to entering on multiple occasions, this thesis is concerned with multiple entry where students enter at least twice before the end of Year 11 and does not consider those students that enter at the end of Year 11 and re-sit in post-16 education, since the drivers for each approach are likely to differ. Moreover, entering at the end of Year 11 and re-sitting in post-16 education is not unusual or unexpected.

At this point it is worth briefly considering the relationship between multiple entry and re-sitting, since there is clear overlap between the two. Indeed, multiple entry is a form of re-sitting, although there are some subtle differences between the re-sitting that has been the focus of the existing literature and the way that multiple entry is defined here. Much of the literature relating to re-sitting in the context of GCSEs (or A levels) has been concerned with the effects of re-sitting individual modules of a modular qualification (see Poon Scott, 2012; Vidal Rodeiro and Nádas, 2012). While the motivation to do this – to improve the final grade – may be similar to the motivation for multiple entry to the whole qualification, students re-sitting modules typically only certificate once (though they may, of course, re-sit the entire qualification). Thus, re-sitting modules is facilitated by the

[22] Note that in modular assessments students can (and often do) re-sit parts of the assessment. This thesis is concerned with instances of multiple certification though, and not re-sitting individual modules.

[23] More recently, JCQ rules have prevented schools from entering students to multiple qualifications in the same examination series, since to do so would require schools to move the start time of the examinations which is not allowed (assessments in the same subject are scheduled at the same time; see JCQ, 2017). It is not clear when these rules came into effect, but it is likely to have been after the analyses reported in this thesis were conducted.

structure of the qualifications. Nonetheless, there are clear overlaps with multiple entry and some of the consequences of re-sitting individual modules of a qualification, such as the effects on student learning and outcomes, may also be relevant to multiple certifications. This is discussed further in Chapter 2.

The other common use of re-sitting is in post-16 education for those students that did not achieve a 'good' pass in their first GCSE attempt, particularly in English and mathematics. The motivation for these students to enter on multiple occasions is likely to be different to the students entering multiple times before the end of Year 11. Indeed, for students re-sitting in post-16 education it is likely that the impetus to re-sit is from the individual's need to achieve a given grade to access future education or employment opportunities (rather than the needs of the school), or more recently, because it is a condition of their funding.[24] Thus, the term multiple entry is being used in this thesis to refer to a specific form of re-sitting.

Although the terms early and multiple entry have been defined separately thus far, they are intrinsically linked: early entry allows students multiple opportunities to certificate before the end of Year 11 (the point at which they would typically be expected to certificate). Indeed, unless students were entering twice at the end of Year 11, without early entry, the practice of multiple entry (as defined here) would be limited. As such, this thesis will focus on EMEE when students enter early in Year 10 or Year 11, and on multiple occasions before the end of Year 11. Having now defined the key terms, the rationale and focus of the research are now considered.

[24] From 2014 students aged 16 – 18 who have not achieved grade C or above in GCSE English and (or) mathematics must continue studying these subjects as a condition of funding. Students achieving a grade D must be enrolled on a GCSE course, providing another impetus to re-sit (see Education and Skills Funding Agency, 2017).

Rationale

The focus of this thesis on EMEE, and specifically on the implications for examination standards, primarily stemmed from two sources: the educational assessment context within which the work in this thesis was planned, and the employment of the author. The educational assessment context shaped the overall focus of the thesis, so we shall consider this first.

Why early and multiple examination entry?

At the time of developing the ideas for this thesis (in 2010), early examination entry to GCSE assessments was on the rise and was increasingly the subject of attention from the media, the public and subject communities (see Chapter 2). This included questions around why schools were increasingly entering a wide range of students early, and what the possible consequences of this might be. Since widespread early entry was a new phenomenon (previously early entry was reserved for the most able students – see Chapter 2), there was little evidence in the literature. Thus, there was considerable debate and speculation around the use of early examination entry, but no real evidence to draw upon. The debate around early entry was increasingly negative, with concerns that early entry was not in the best interests of students but was enacted for alternative reasons, such as the drive to ensure that students achieved certain grades for school performance measures (see Chapter 2). This growing debate, and the lack of existing evidence, suggested that there was a gap in the literature that this thesis could contribute to.

Although this backdrop provided the main impetus for this thesis, the research also evolved over time. At the point that this thesis was planned much of the public discourse and media attention was focused specifically on early entry and there was less reference to multiple entry. As such, the initial intention was to focus this thesis on early examination entry and to consider some of the potential implications of this approach – for example, how early entry might influence students' performance and attainment. However, as the early chapters of this thesis were progressing, it became increasingly apparent that

multiple examination entry was also key, and that the two approaches were closely linked. As such, studying one approach but not the other would limit the extent to which the wider context that these phenomenon operate in was considered. The focus of this thesis was therefore widened to consider EMEE. The first set of quantitative analyses therefore focus on early entry, while the latter phases focus on EMEE.

While EMEE was increasing in both GCSE English and mathematics between 2007 and 2013, the entry trends in each subject differed (EMEE was more prevalent in mathematics; see Chapter 2), as did the level of public debate, which was generally centred on mathematics. The higher prevalence of EMEE in mathematics is perhaps not surprising given the features of the qualifications in each subject and how these qualifications operated (see Table 2). For example, due to the nature of the assessment, mathematics naturally lent itself more readily to repeated entry than English because the content was broadly similar across exam boards and there was no controlled assessment. Furthermore, there were more opportunities to enter GCSE mathematics since there were four examination series each academic year, and multiple qualifications were available in the same examination series. This suggests that mathematics could be considered the more extreme example of EMEE, and therefore likely to offer the most interesting insight into this approach. It is for this reason, and my own experiences of standard setting (which were primarily related to GCSE mathematics – see later discussion in this chapter), that guided the focus of this thesis to EMEE in mathematics.

Table 2. *Similarities and differences between GCSE mathematics and English (2007-2014)*

	Mathematics	English
Qualifications	2 from each exam board	1 from each exam board
Structure	Entirely externally assessed from 2009	Controlled assessment (60%) and external assessment (40%)
Tiering	Higher and foundation tier from 2008	Higher and foundation tier
Assessment opportunities	January, March, June and November prior to 2013	June and November
Modular or linear	Both modular and linear prior to 2014	Linear prior to 2012, modular 2012 – 2013, linear from 2014
Accountability	Required in five A* – C measure from 2007	Required in five A* – C measure from 2007
Progression	Grade C often required to access further education or employment	Grade C often required to access further education or employment

Having identified a gap in the literature, a secondary concern was where the specific focus of this thesis should lie. Given that there was limited academic literature referring to EMEE, there were many possible avenues to pursue. For example, one might consider how entry trends were changing, why students were being entered early or on multiple occasions, and the possible consequences of these approaches. While the former two questions seem key to understanding the phenomenon and were therefore two focal points for this thesis, considering the consequences of EMEE is more challenging. This is because there are likely to be multiple consequences of any phenomenon for the various stakeholders involved – in this case students, teachers, schools, and exam boards – and studying all of these would be beyond the scope of one thesis. For example, much of the public discourse around EMEE focused on the potential consequences for students, such as the impact on their performance, motivation, and future progression. However, there are also other possible consequences – for example, for the uptake of particular subjects

in post-compulsory education. Ultimately, it was the employment of the author at the time that shaped the focus of this research.

Why standards?

At the time that this thesis was planned I was working as a researcher for the exam board, AQA, and had done so for a number of years. As part of my role I was heavily involved in supporting the setting of grade boundaries, in particular for GCSE mathematics. I was therefore very familiar with the approach to maintaining standards (outlined earlier in this chapter), and had developed considerable technical knowledge and expertise in relation to this during my employment. Such depth of understanding is difficult to develop without being embedded in the process, but would be invaluable for shaping the work in this thesis.

As part of my employment I was therefore observing first-hand the rise in early entry over time, since I had access to data about the types of students that were entering each qualification in each examination series. I also had shared responsibility for ensuring that despite such changes in entries, examination standards were maintained over time. It is this background that prompted me to focus the thesis on the implications of EMEE for maintaining examination standards, since I was in the fairly unique position of having the understanding and knowledge about standard setting that could be applied to this important research topic.

In determining the focus of this thesis I was also aware that while considerable expertise is held within the exam boards in relation to the operational aspects of maintaining standards, little of this operational work is documented externally or in the academic literature. This thesis therefore sought to bring my own operational experiences and knowledge to the issue of EMEE, and generate evidence that could sit within the limited literature that currently exists. Working in an awarding body for a number of years prior to starting this work has afforded me with knowledge and insight that it would be difficult to gain without such experience, which I feel has put me in a fairly unique position for

studying the phenomenon in this thesis (the potential implications and influences of the author are discussed in Chapter 3). Furthermore, working in an awarding body has allowed me access to data that external researchers might not even be aware of. I am therefore grateful to the sponsors of my research and the individuals at these institutions, for allowing me access to data and information and for supporting me to study part-time alongside my employment – AQA who supported me through the first four years of this work, and Ofqual who supported me throughout the remainder of this work.

While the focus of this thesis on the implications of EMEE for maintaining standards was prompted by the author's operational experiences and responsibilities, any consideration of examination standards must also grapple with the question of how examination standards are, or should be, defined. This is a contentious area that continues to be debated (see Chapter 8). Thus, when considering the implications of EMEE for examination standards, this includes considering how examinations are maintained in practice (from an operational perspective), but also how examinations standards are defined. The theoretical literature relating to how examinations standards are defined thus provides the theoretical framework within which the implications of EMEE are considered.

A final point to consider is that when studying any topic within the field of education and assessment the policy backdrop is likely to change and evolve as time moves on. This is particularly the case for the research in this thesis, since the work was conducted part-time and therefore spanned a longer time period – between 2011 and 2017. As such, the work has progressed against a backdrop of policy changes – most importantly the changes to accountability measures announced by the Government in September 2013 (DfE, 2013a) that signalled the return of single examination entry for the majority of students. While this research has been able to study the practice of EMEE as the situation unfolded, this has also had implications for the way in which the work evolved. For example, the initial intention of this thesis – to focus on early examination entry – evolved to incorporate a consideration of multiple entry, as the intrinsic link between these two

approaches became apparent. The thesis is thus presented as a series of studies that were conducted at particular points in time, and no attempts are made to update earlier analyses based on the most recent data available. To do otherwise would risk the story of EMEE being lost. The thesis therefore presents the situation as it occurred in practice and captures the story of EMEE as it unfolded.

Contributions

The work presented in this thesis draws on a number of areas of literature. This includes the existing literature relating to EMEE (much of which resides in the grey literature – see Chapter 2), the literature relating to how examination standards are maintained, and the literature relating to how examination standards are defined (see Chapter 8). Similarly, the work is considered to contribute to a number of areas. Primarily, the work will add to the literature relating to EMEE which at the start of this thesis was, and indeed remains, relatively limited. This is perhaps not surprising since widespread EMEE was limited to the period between 2007 and 2013. Thus, when changes to accountability measures were announced in 2013 that effectively ceased EMEE, research attention was likely to have moved elsewhere. Nonetheless, this is not considered to detract from the work presented in this thesis and the contributions that this research can make. There are two main reasons for this.

First, given the cyclical nature of policy changes and reforms (Baird and Hopfenbeck, 2016), many of the issues and policies identified in this thesis are likely to recur at another time, or indeed in another place. As an example, at the time that this thesis was nearing completion EMEE to GCSE assessments had become an issue in Wales. This prompted the examinations regulator (Qualification Wales) to launch a research study exploring schools' approaches to EMEE (Qualifications Wales, 2017; see Chapter 2), that in turn resulted in the Welsh Government announcing changes to accountability measures to discourage EMEE (Welsh Government, 2017). Thus, the more recent situation in Wales

draws parallels with the phenomenon presented in this thesis, and many of the challenges that faced the assessment system in Wales are the same as those considered here.

A second reason that this thesis is considered to be of more than just historical interest is that the findings are likely to have wider importance beyond the specific policy identified here, in particular, in relation to the influence of performance measures on schools. A large body of literature currently exists relating to the way in which performance measures can interact with and influence behaviour in schools (see Acquah, 2012, for an overview and Chapter 2). Thus, the findings in this thesis are considered to have relevance for the way in which performance measures can influence behaviour within schools that transcends EMEE. In addition to adding to the literature on EMEE, the work can contribute to the body of literature that considers the interaction between performance measures and schools' behaviour.

Summary of Chapter 1

This chapter has provided the context to allow readers to engage with the work presented in this thesis. This has included an overview of the assessment system in England, focusing on the GCSE qualification, and a discussion of the methods used to maintain standards. The chapter has also provided the rationale for this research – to consider the use and drivers of EEME to GCSE mathematics assessment and the implications for examination standards – and the contributions that the research makes. The chapter has also discussed how the research has evolved in response to policy changes that have occurred as the research progressed. Definitions of the key terms in this thesis have also been provided.

The following chapter discusses the literature relating to EMEE and provides an overview of the key research questions considered in this thesis. Further literature is introduced at relevant points later in this thesis – in Chapter 6 there is a consideration of the literature

relating to assessment unreliability, and in Chapter 8 there is a consideration of the literature relating to how examination standards are defined.

Chapter 2 – Literature review

Overview

This chapter provides an overview of the literature relating to EMEE. The chapter considers the origins of early entry for gifted and talented students, the subsequent proliferation of early entry for a wider range of students, and the emerging link between early entry and multiple entry. While early entry for the most able students differs from EMEE as defined in this thesis (see Chapter 1), understanding the historical background is important when considering the context within which EMEE rose in prominence. Furthermore, some of the issues relating to early entry for the most able overlap with those relating to EMEE for a wider range of students.

The chapter also discusses the possible drivers of EMEE and the potential effects for students, schools and the wider assessment system. Finally, the chapter discusses the limitations of the existing literature and outlines the key research questions addressed in this thesis. The literature review is written to reflect the manner in which this thesis evolved (see Chapter 1) – focusing initially on the increase in early entry and the subsequent proliferation to EMEE. The literature review therefore focuses on the rise in EMEE and not the subsequent policy changes that effectively ended the widespread use of EMEE (this is discussed in Chapter 8).

Much of the research presented in this chapter resides in the grey literature since there is little academic literature relating specifically to EMEE. Additionally, much of the literature was published around the time that this thesis began and there is less recent evidence to draw upon. This is perhaps not surprising given that EMEE in England effectively ceased following Government changes to accountability measures in 2013 (DfE, 2013a), meaning that research attention is likely to have focused elsewhere. Nonetheless, there is more recent evidence from other jurisdictions, in particular from Wales, that is discussed at the end of this chapter.

The history of early and multiple entry

Early entry for the most able…

Despite becoming more prominent in recent years and being linked with multiple entry, the concept of early entry in itself is not a new one. A study from 1980, for example, exploring school curriculum and assessment standards highlights the issue of early entry which, in this case, was reserved for the more able students who entered English and mathematics O-levels early, a policy adopted more frequently by grammar schools (Bryan and Digby, 1983; see also Bell, 2001). Other sources concur that in the past early entry was largely used to accelerate the more able students (DfE, 2011; Pope and Noyes, 2011), a practice that was likely to have been encouraged by an increased focus on providing for gifted and talented students in recent decades (Department for Children, Schools and Families, 2007; Department for Education and Employment, 1997; DfES, 2001, 2005b; Freeman, 2002). For example, the 2001 Government White Paper 'Schools Achieving Success' (DfES, 2001) set out a policy for gifted and talented children that aimed to,

> *"Blend increased pace, depth and breadth in varying proportions according to the ability and needs of pupils. We want teachers to consider express sets, fast-tracking and more early entry to GCSE and advanced qualifications"*

The White Paper also stated the need to re-consider the assumption that students should progress through the education system at the same rate and outlined the Government's intentions to amend legislation to allow more students to take examinations early. The need to provide for gifted and talented students continued to feature in later White Papers (DfES, 2001; 2005b), resulting in a number of approaches being adopted by schools; this included early entry to facilitate acceleration and enrichment.

Acceleration and enrichment

The desire for increased provision for the most able students, including early entry, opened the debate regarding the use of acceleration and enrichment strategies. In the UK context, acceleration has typically been used to refer to two scenarios: instances where a group of students sit qualifications early and progress to more demanding work ahead of their peers, and situations where individual students work with older students for certain parts of their timetable (Westminster Institute of Education, 2006). Although the meaning can be slightly different in the United States (US) context (where students can skip entire academic years), it is here where much of the evidence concerning the effects of acceleration resides. Acceleration has been well researched over recent years and the observed effects are consistent within the literature (e.g. see Colangelo, Assouline, and Gross, 2004, for a review; Reis and Renzulli, 2010). Such effects include positive benefits for students' attainment and future progression, as well as benefits for their academic experience and motivation.

The use of acceleration strategies in England, including early entry, was undoubtedly encouraged by the Government in recent years (DfES, 2001; 2005b) and can evidently be beneficial (see Colangelo, Assouline, and Gross, 2004). However, this model is not without criticism, and some of these criticisms overlap with the issues associated with EMEE (as discussed later in this chapter). For mathematics, critics argue that acceleration (even for the most able) can have disadvantages (UK Mathematics Foundation, 2000, cited by Koshy and Casey, 2005). For example, it can result in students not learning more mathematics but learning the same amount sooner (Fielker, 1997, cited by Koshy and Casey, 2005), meaning that there are no benefits in terms of students' learning.

In relation to GCSE assessments, acceleration for the more able students is considered particularly problematic when schools do not have sufficient provision for students who complete a qualification early (see Noyes and Sealey, 2011), with many organisations arguing that acceleration should only be adopted when there is suitable provision for

students the following year (e.g. The Institute of Mathematics and its Applications and the London Mathematical Society [IMA and LMA], 2011). This provision could involve an enriched curriculum being offered (e.g. allowing students to study additional qualifications in Year 11; see Noyes and Sealey, 2011), or allowing students to start the A-level programme of study early, taking this over three (rather than the usual two) years (Bond, Green and Jaworski, 2009) to aid the transition from GCSE to A level (Mathematics in Education and Industry, 2009). Some of these issues will be returned to later when EMEE are considered.

While there is some debate over the merits of acceleration and early entry for the most able students, the use of enrichment is more widely supported. Enrichment typically involves students studying a subject in greater depth or breadth and can be used alongside acceleration (i.e. where students enter early and take additional qualifications the following year), or in isolation (i.e. where students study additional material or qualifications but do not sit assessments early). There is considerable support for enrichment as a means of providing for the most able students, particularly in the mathematics community where enrichment is frequently cited as a preferable strategy to acceleration (e.g. The Mathematical Association, 2010; Winter, 2001). Moreover, there is evidence that many in the teaching profession historically shared this view (Chyriwsky and Kennard, 1997). In a survey of 524 mathematics teachers enrichment was cited as the preferred strategy for the most able children, with fewer than half of respondents supporting early entry (46.6%; Chyriwsky and Kennard, 1997). While the debate regarding acceleration and enrichment for the most able students will no doubt continue, at GCSE it was the increase in EMEE for students with average and lower ability that began to attract more interest from stakeholders in the assessment system. This included the DfE, subject associations, and the media.

The rise in early entry

While early entry at GCSE was traditionally reserved for a small group of highly able students (despite some debate regarding its merits), between 2007 and 2010 there were significant increases in the proportion of students entering early. One of the earliest pieces of evidence showing this rise in early entry was a report by the DfE (2011) that aimed to explore the impact of early entry on performance, with a view to informing stakeholders of the possible effects (see later discussion in this chapter). Focusing on GCSE English and mathematics, the report highlighted the increase in early entry between 2007 and 2010: while around two per cent of the GCSE English and five per cent of the GCSE mathematics national cohorts in England were early entries in 2007 (across all examination series), this figure had risen to around 25% by 2010.

A number of other sources reported similar findings.[25] A report by Cambridge Assessment that explored the age of GCSE students showed that the percentage of early entries in English and mathematics increased between 2004 and 2009 (Gill, 2010), while a subsequent report showed that the percentage of GCSE mathematics entries from 15-year-olds continued to rise between 2011 and 2013 (Zanini, 2015). Other findings showing an increase in early entry during the same time period, particularly in English and mathematics, were reported elsewhere (e.g. see Advisory Committee on Mathematics Education [ACME], 2011; Gill, 2012, 2013; Ofsted, 2013; Ofqual, 2013b).

Although there is evidence of early entry being used prior to 2007, the sharp increase in early entry appears to have mainly occurred between 2007 and 2010 (DfE, 2011). As shown by Gill (2010), the percentage of GCSE entries across all subjects from 15-year-old students remained relatively stable between 2000 and 2004 (at 1.2% and 1.7% of the total

[25] Note that the parameters of these analyses generally differed, so the figures are not directly comparable across different reports. This is also true for those reports focusing on multiple entry, since the definition of multiple entry was not necessarily the same.

entries, respectively), yet increased dramatically in the subsequent five years to reach 5.1% of the total entries in 2009. During this time, the number of 15-year-old GCSE students as a percentage of the overall number of students also increased, from 7.5% to 19.2%. The greater percentage of 15-year-old students, relative to the percentage of 15-year-old entries, suggests that the younger students were only entering a small number of GCSE subjects early. Indeed, as discussed by Gill (2010), there is evidence that the number of GCSE subjects that younger students took remained stable over time, suggesting that schools were only using early entry for certain subjects.

In addition to the increase in 15-year-old students certificating at GCSE it is worth noting that the number of even younger students sitting GCSEs also increased around the same time (see Gill, 2010). However, these numbers remained relatively small and were likely to represent unusual situations with specific motives, so are not the focus of this thesis (see Chapter 1).

Early entry for a wider range of students

The increase in early entry was particularly prominent in GCSE English and mathematics (ACME, 2011; DfE, 2011; Gill, 2010, 2013; Ofqual, 2013b; Pope and Noyes, 2011), with an increasing number of schools reporting that they used early entry for a wider range of students than just the most able. For example, a number of National Challenge[26] schools reported entering 90% or more of their students early for GCSE mathematics in 2009 (ACME, 2011; see also DfE, 2011), while in a survey administered by Ofsted (2013) a number of schools reported entering 99% or more of their students early in 2011. Furthermore, a survey undertaken in summer 2011 with a representative sample of 692 secondary schools in England revealed that 81% of respondents entered some pupils in

[26] The National Challenge aimed to ensure that at least 30% of students in all secondary schools achieved five GCSEs at grades A* to C, including English and mathematics. National Challenge schools were identified as those that fell below this target (Balls, 2008a).

Years 9 and 10 for some GCSEs, with the majority reporting no plans to change their approach at that time (Centre for Analysis of Youth Transitions, 2011).

The increase in early entry was coupled with a decline in outcomes for early entrants (DfE, 2011; Gill, 2013) suggesting that, as a cohort, students who entered early were performing worse. For example, the DfE (2011) showed that the cumulative percentage outcomes for early entrants in GCSE mathematics declined between 2008 and 2010, suggesting a shift in the ability profile of students who were entering early, or, a change in the performance of younger students. Similarly, Gill (2013) showed that between 2009 and 2011 the percentage of students that entered GCSE English or mathematics early and achieved a grade A on their first attempt generally declined (although this also depended to some extent on when students entered early).

Alongside this decrease in outcomes there was also evidence that students who entered early tended to perform worse than those entering at the end of Year 11 in terms of achieving the highest grades of A* – B (DfE, 2011). Since this comparison did not account for the prior attainment of the students it is not possible to draw any conclusions relating to how early entry might influence an individual student's performance (this is discussed later in this chapter). However, this further suggests that schools were increasingly entering a wider range of students early than just the most able (see also Ofsted, 2012).

There was also evidence of trends at the school level: lower attaining schools tended to be increasingly entering students early (DfE, 2011; see also Ofsted, 2013), with independent schools the least likely to adopt this approach (DfE, 2011; Gill, 2012; 2013; Noyes, Drake, Wake and Murphy, 2010). For example, while 11% of academies or comprehensive schools entered no students early for GCSE English and mathematics in June 2010, this figure rose to 58% for independent schools (DfE, 2011). Similarly, Zanini (2015) found that in 2013 the percentage of GCSE entries that were from 15-year-old students in independent schools was lower than in comprehensive schools or academies, a finding observed in previous years (see Gill, 2012).

In addition to these trends by school type, there was also evidence that the proportion of early entries was greater amongst those schools that had a higher percentage of students eligible for free school meals (FSM; DfE, 2011; Ofsted, 2013). Given that, on average, FSM students tend to perform worse than non-FSM students (DfE, 2011), this again suggests that a wider range of students were being entered early than before.

Early entry coupled with multiple entry

The majority of the reports discussed above focused on early entry. This is perhaps not surprising given that much of the debate relating to EMEE initially focused on the rise in early entry and the possible negative consequences of this approach. Thus, the link between early and multiple entry emerged later (see Chapter 1). Nonetheless, the DfE (2011) reported some figures relating to multiple entry in their early report: in 2010 less than 20% of students entered GCSE mathematics more than once before the end of KS4, with the majority of multiple entrants entering twice.[27] While the percentage of students entering on multiple occasions was therefore relatively low, multiple entrants still represented around 100,000 students, suggesting that the prevalence of multiple entry was also likely to be increasing during this period.

The DfE (2011) report also discussed the issue of re-sitting and highlighted the link between early entry and the increased chance of students re-sitting (see Chapter 1 for a discussion about multiple entry and re-sitting). For example, of those students entering GCSE mathematics early in 2010, 68% re-took in a later examination series, with 45% improving their final grade. As would be expected, a greater proportion of those achieving the lower grades re-sat both GCSE English and mathematics, regardless of when their initial entry was made (i.e. at the end of Year 10 or mid-way through Year 11), with re-

[27] Note that no figures relating to multiple entry were reported for GCSE English or any other subjects. It is not clear whether this is because the numbers were small or whether the analyses were not conducted.

sitting concentrated around the grade C/D borderline. For example, 98% of students achieving a grade D in GCSE mathematics at either the end of Year 10 or mid-way through Year 11 re-took in a subsequent examination series (DfE, 2011). This further suggests a link between early and multiple entry, though it is perhaps not clear at this point which is driving which: whether early entry is used to allow multiple entry, or whether early entry is key and multiple entry follows depending on the outcomes in a student's first attempt. It is likely that they interact though, depending on an individual student's circumstances.

Although there was limited evidence reported by the DfE (2011) related to multiple entry, other publications considered this in more detail in subsequent years, arguably when the link between early and multiple entry became more apparent (see Chapter 1). The first of these was a short report published by Ofqual (2013c). Focusing on the summer 2012 examination series,[28] this report showed that multiple entry was far more prevalent in mathematics than English, with multiple entrants accounting for around 15% of all students certificating in GCSE mathematics.[29] Furthermore, the proportion of students achieving the top grades (A* – B) was higher for the single entrants compared to the multiple entrants. This suggests that multiple entry was more prevalent for the lower ability students.

Another report that considered the presence of multiple entry was published by the DfE shortly afterwards (DfE, 2013b). This report focused on multiple entry in both English and mathematics for students reaching the end of Year 11 in summer 2011 and 2012. The analyses showed that while the percentage of students that entered on multiple occasions

[28] Note that this analysis also included entries prior to summer 2012 for those students also entering in summer 2012.

[29] Note that the definition of multiple entry differed to the definition used in this thesis, since entries to units of the modular qualifications in the two years prior to 2012 were also included in the analyses.

in English declined between 2011 and 2012 (from 25% to 7%), in mathematics they rose considerably, from 28% in 2011 to 41% in 2012.[30] The subsequent year the percentage of multiple entrants increased in both subjects though, as shown in a statistical first release published by the DfE (2013c): thus, by summer 2013, almost half of students reaching the end of Year 11 had entered GCSE mathematics more than once.

Returning to the 2011 and 2012 data, further analyses showed that the increase in multiple entry in mathematics was observed across the board in terms of the number of entries each student made: the percentage of students entering twice increased from 23% to 26%, the percentage of students entering on three occasions increased from 5% to 11%, and the percentage of students entering on more than three occasions increased from 1% to 4% (DfE, 2013b). The analyses also showed that the percentage of multiple entrants was higher in sponsored academies[31] than similar matched schools, for both English and mathematics, suggesting that early entry was more common in some school types than others.

The analyses also considered whether multiple entrants entered the same or different specifications[32] for each entry attempt (DfE, 2013b). In English, in summer 2012 the majority of students (92%) that entered twice sat the same specification on both occasions, while the majority of students (80%) that entered on three occasions entered different specifications. For mathematics, entry to different specifications was more common: of the students entering twice 47% sat different specifications on each occasion, while this rose to 78% for students that entered three times. The approach of entering

---

30 Note that a different definition of multiple entry was used here to the Ofqual report, hence the difference in the percentage of multiple entrants.

31 A sponsored academy is a former maintained school that has been transformed to academy status as part of Government intervention. Such schools are likely to have been lower performing prior to their conversion to an academy.

32 The term specification is used throughout this thesis to refer to different versions of the same qualification, e.g. awarding bodies might develop multiple specifications for the same GCSE subject.

multiple specifications in mathematics also rose between 2011 and 2012 (only 14% of the students that entered twice sat multiple specifications in summer 2011).

The wider use of multiple specifications in mathematics compared to English is perhaps not surprising since there has been no internal assessment in mathematics since 2009 (see Chapter 1). Thus, students would not be required to take multiple versions of the internal assessment in mathematics as they would in English. However, the approach of entering students to multiple specifications is perhaps also an indication that schools were using EMEE for other reasons than to benefit the students: it is difficult to conceive how requiring students to complete multiple internal assessments (in English) or adapt to different styles of examination papers (in English and mathematics) would be beneficial.

Another more recent report considering multiple entry was published by Cambridge Assessment (Vidal Rodeiro, 2014). This report reviewed the extent of multiple entry in the same summer examination series for the academic years ending in 2012 and 2013 (Vidal Rodeiro, 2014). While this form of multiple entry is far more specific than the definition of multiple used in this thesis (see Chapter 1), the analyses provide further insight into multiple entry and the type of students entering on multiple occasions.

Considering both GCSEs and international GCSEs,[33] Vidal Rodeiro (2014) found that the number of students entering multiple times in the same summer examination series was higher in English and mathematics than other subjects in both 2012 and 2013.[34] There were also increases in the percentage of multiple entries between the two years for both subjects. In 2012 the percentage of students entering multiple times in the same

[33] International GCSEs (also known as level 1/2 certificates) were alternative qualifications to GCSEs offered by the main UK awarding bodies (see Chapter 1) and Cambridge International Examinations (CIE). They rose in popularity from around 2010, with entries peaking in 2016. Following the latest round of GCSE reform (for subjects first assessed in 2017), international GCSEs were not included in performance tables and the entries dropped significantly (Ofqual, 2017d).

[34] For other subjects the percentage of students entering multiple times was less than 0.2% each year.

examination series was 0.32% in English language and 3.75% in mathematics, which rose to 4.18% and 4.69%, respectively, in 2013.

Of the multiple entrants in mathematics, the majority entered two specifications, although there was also evidence of a small number of students (303 students or 0.07% of the total entry in 2012, and 275 students or 0.08% of the total entry in 2013) entering three specifications. In English there were only two students entering three specifications in 2013 (in 2012 all of the multiple entrants entered two qualifications). The higher number of students entering more than two specifications in mathematics compared to English may again be due to the presence of internal assessment in English, since to complete multiple pieces of internal assessment for the same examination series would increase the burden on students and schools.

In addition to showing the percentage of multiple entries each year, the report also considered the type of students that entered on multiple occasions. The analyses showed that compared to the single entry students, multiple entrants were more likely to be from state-maintained schools (academies or comprehensives) than independent or selective schools (Vidal Rodeiro, 2014). Furthermore, multiple entrants were more likely to have lower prior attainment and higher levels of deprivation, were more likely to perform worse at GCSE than the single entrants, and were typically concentrated around the grade C/D borderline. This suggests that multiple entry in the same examination series was being used for a certain type of student, generally the lower ability students.

Summary of changes in entry trends

To summarise, a number of reports considering the prevalence of early and multiple entry were published around the time that this research was planned. These reports each adopted a slightly different approach to their analyses, meaning that the figures varied. Nonetheless, there are clear overall trends: while early entry was traditionally reserved for a small number of the more able students, there were significant increases in the

proportion of early entries from around 2007, particularly in GCSE English and mathematics. These increases were coupled with a decline in the outcomes that early entry students achieved, suggesting that a wider range of students were entering early, rather than just the most able. Furthermore, early entry was particularly prevalent in lower performing schools and schools with a higher proportion of students eligible for FSM, and less prevalent in independent schools. As the prevalence of early entry increased, the link between early and multiple entry also emerged and there was evidence of increased multiple entry, particularly in GCSE mathematics. Finally, compared to single entrants, multiple entrants tended to have lower prior attainment and higher levels of deprivation, be concentrated around the grade C/D borderline, and were less likely than single entrants to achieve the top grades.

The following sections considers why there was a rise in EMEE during this period, by drawing on the various policy changes that were being implemented and other possible drivers. The policy decisions and their perceived impact on EMEE are considered within the context when EMEE was increasing, rather than at the present time. This is to ensure that the context within which the work presented in this thesis developed is provided.

Why more early entry and multiple entry?

The factors influencing why students were increasingly entered early are likely to be complex and inter-linked. While many sources were critical of EMEE and cited policy decisions at Government level as the main driver (e.g. see ACME, 2011; Perryman, Ball, Maguire and Braun, 2011; The Mathematical Association, 2010), there were also likely to be instances of schools using some form of EMEE to benefit their students. The policy changes that were thought to have influenced schools' entry practices related to decisions concerning the structure and format of the assessment system, as well as those related to accountability and performance measures, the latter of which were likely to have different implications for different subjects. For example, initiatives such as performance league

tables and the introduction of the EBacc (see Chapter 1) were likely to have placed increased emphasis on certain 'key' subjects.

While much of the debate relating to the factors that might have influenced the rise in EMEE was based on anecdote, one study was published by Ofsted that attempted to gain further insight into schools' entry approaches (Ofsted, 2013). This study was published in early 2013 (prior to changes to accountability measures that were announced later that year; see Chapter 1) and sought to investigate why schools were using EMEE, which schools were using EMEE, and the possible effects on attainment (see later discussion in this chapter). While this report came quite late in the story of EMEE, it is nonetheless a useful piece of evidence to draw upon as we consider the possible factors that influenced the rise in EMEE.

The following section is presented chronologically so that the policies thought to have influenced the rise in EMEE (as it was happening) are discussed first, followed by the later evidence that emerged. The number and pace of policy changes in education has seemingly increased dramatically over recent years (Braun, Maguire and Ball, 2010), alongside increased Government influence over the qualification system (Isaacs, 2010). This often makes it difficult to fully understand and isolate the impact of each policy decision on a particular issue such as EMEE. Nonetheless, the following section outlines some of the policy developments relating to GCSEs and discusses how such changes might have impact upon EMEE.

Key policies

School performance and accountability measures

One of the most significant initiatives thought to have influenced EMEE is the use of performance measures for school accountability purposes (see Chapter 1). While these measures have changed over time (and continue to do so; see Chapter 1), during the period that this thesis relates to the key measure was the percentage of students (within

each school) that achieved five GCSEs at grades A* – C, including English and mathematics.

The nature of this measure, in particular the focus on English and mathematics, resulted in grade C becoming an important threshold at GCSE, thereby increasing the pressure on schools (Gillborn and Youdell, 2000; Noyes and Sealey, 2011; Perryman, Ball, Maguire and Braun, 2011; Pope and Noyes, 2011). This is considered to have prompted the use of a number of pedagogical approaches within schools to improve results, many of which are documented in the literature. This includes focusing teaching resources upon borderline students, the scheduling of intensive revision sessions, setting students by ability, and 'teaching to the test' (Ball, Maguire, Braun, Perryman, and Hoskins, 2012; Burgess, Propper, Slater, and Wilson, 2005; Gillborn and Youdell, 2000; Goldstein, 2004; Hutchings, 2015; Noyes, Drake, Wake, and Murphy, 2010; Perryman, Ball, Maguire and Braun, 2011; Stobart, 2008; West, 2010; Wilson, Croxson and Atkinson, 2006; Wilson and Piebalga, 2008). Such behaviours are not restricted to the English assessment system (see Jacob and Levitt, 2003, and Reback, 2008, for evidence from the US) yet are unlikely to be without consequence: there is evidence that focusing on borderline students can be detrimental to the attainment of other students (Burgess, Propper, Slater and Wilson, 2005; Gillborn and Youdell, 2000), that 'teaching to the test' can distort the curriculum (Goldstein, 2004), and that focusing on particular subjects can narrow the curriculum that students experience (Hutchings, 2015).

The existing literature also suggests that focusing on the grade C threshold is likely to have influenced schools' entry approaches to assessments, though this evidence is more limited. Nonetheless, there is evidence of students being entered for alternative assessments to GCSEs that were rewarded highly in performance tables (Hutchings, 2015; Perryman, Ball, Maguire and Braun, 2011; Stobart, 2008; Wolf, 2011) or perceived as 'easier' (Burgess, Propper, Slater and Wilson, 2005; West, 2010), of students not being entered if they are considered unlikely to achieve the required threshold (de Wolf and

Janssens, 2007), and of schools strategically entering students to tiered assessments (see Chapter 1; Gillborn and Youdell, 2000; Strand, 2012).

The presence of performance measures was also thought to be influential in the rise in EMEE. The focus on the grade C threshold was thought to have led schools to become pre-occupied with examination results, prompting the use of an increasing number of strategies to try and improve results, including early entry to give students the opportunity to re-sit if necessary (Perryman, Ball, Maguire and Braun, 2011). This view was supported by several sources. In 2011, ACME stated that they believed "*an all-consuming focus on league tables is skewing behaviour in schools"*, forcing schools into prioritising performance measures over the interests of individual students (ACME, 2011). This view was shared by The Mathematical Association (2010), who cited government policy and the resultant fears of head teachers, teachers and parents as a key driver of EMEE. Thus, it was argued that, as a reaction to increased accountability pressures, schools used early entry to give students multiple attempts to achieve a grade C, particularly in English and mathematics.

The introduction of the English Baccalaureate

A further policy change that could have influenced EMEE was the introduction of the EBacc, announced in The Schools White Paper in 2010 (DfE, 2010; see Chapter 1). Although there was uncertainty about the possible influence of the EBacc on EMEE as it was introduced, there were some indications that it could influence schools subject choices and possibly their entry policies to GCSE, by putting additional pressure on students' performance in subjects other than English and mathematics (Perryman, Ball, Maguire and Braun, 2011). For example, in a survey of a representative sample of 692 secondary schools in England undertaken between June and July 2011, around half of respondents reported that the EBacc had influenced the curriculum that they intended to offer (Centre for Analysis of Youth Transitions, 2011). Furthermore, within these schools an increasing proportion of younger students were taking EBacc subjects, suggesting that

its introduction has influenced students' subject choices (although this may have also been influenced by schools changing the subjects that they offered to students).

Other sources cite a similar effect of the EBacc, with institutions changing their curriculum offerings to focus on EBacc subjects (Hodgson and Spours, 2011) and the popularity of certain non-EBacc subjects declining (Mansell, 2011), a phenomenon that has continued in more recent years (e.g. see Ofqual, 2017d). While this suggests that the EBacc was influencing schools' curriculum provision, there was less evidence that the EBacc influenced EMEE. For example, surveys by Ipsos Mori (2013) on behalf of the DfE and the Centre for Analysis of Youth Transitions (2011) suggested that only a small number of schools were planning to change their early entry policies in response to the introduction of the EBacc. Nonetheless, the EBacc was another policy change that schools had to react to around the time that EMEE was increasing.

The removal of KS3 tests

In addition to policy changes relating to performance measures and accountability, changes to the structure of GCSEs and the wider assessment system were also likely to have influenced the rise in EMEE. The potential for entering students early to GCSE was likely to have been encouraged following the decision in October 2008 by the then secretary of State for Education, Ed Balls, to abolish KS3 testing for 14-year-olds (Balls, 2008b), national tests that had been compulsory since 1993 at the end of Year 9 (age 11). When announced, the removal of KS3 tests attracted widespread support from a number of sources including teaching unions, parents and other organisations (e.g. see Lipsett, 2008; Paton and Moore, 2008). However, concerns were subsequently expressed regarding how this change might impact upon schools' entry policies at GCSE. A number of sources suggested that the removal of KS3 tests was likely to have encouraged schools to begin the KS4 programme of study early and enter students early for GCSE assessments (e.g. ACME, 2011; Noyes, Drake, Wake and Murphy, 2010; Pope and Noyes, 2011). In some respects, the removal of KS3 tests might have been beneficial if

students were not distracted by assessments at the end of KS3 so would have more time to engage with the GCSE subject content than before. However, if this was coupled with early entry to GCSE and insufficient subsequent provision for students there may also have been detrimental effects, an issue that is discussed later in this chapter.

Modular assessments at GCSE

The removal of KS3 tests was shortly followed by changes to the structure of the GCSE qualification for the majority of subjects first taught in 2009, when most of the newly reformed GCSEs adopted a modular structure (see Chapter 1). Although schools could still adopt a linear approach to entering students to the assessments, many took advantage of the new structure by entering students for modules early (during Years 9 and 10), thus providing opportunities for students to re-sit modules throughout the remainder of the course and improve their final grade (Spalding, 2009; Vidal Rodeiro and Nádas, 2012).[35]

While there is no evidence to suggest that modular assessments in themselves encouraged students to certificate at GCSE early or on multiple occasions they inevitably, by design, encouraged students to sit some GCSE assessments (or modules) prior to the end of Year 11, resulting in a greater number of younger students (Year 9 and 10 students) being assessed at GCSE level. Given that students sitting modules early were still expected to demonstrate the same level of performance as students sitting the assessments at the end of the course (i.e. no allowances were made for younger students), this raised the issue of how maturation might affect a students' performance. It was argued that students sitting modules early might lack the maturity to perform to their full potential, even if they have studied the whole programme of study (Clarke, 1996; Taverner and Wright, 1997; Vidal Rodeiro and Nádas, 2012).

[35] In the modular GCSE structure students were allowed to re-sit each module once, with the highest mark counting towards certification (QCA, 2008).

The availability of modular assessments was short-lived since the coalition Government implemented a return to linear examinations in all GCSE subjects for first teaching from September 2012 (see Chapter 1). Nonetheless, it is likely that the availability of modular assessments will have had some influence on schools' entry policies, possibly encouraging EMEE to modular assessments or certification.

The motives of schools

The policy changes outlined above were thought to have influenced how schools organised their curriculum time and their approaches to entering students to assessments, including in relation to EMEE. In addition, there were also likely to be decisions and motives within individual schools that influenced EMEE, although much of the evidence concerning this was anecdotal. For example, there were reports that one motive for entering students early was to raise the profile of schools or as an example of a school's gifted and talented provision (ACME, 2011; Winter, 2001). Indeed, a quick search of the internet at the time this work began revealed a number of schools that published documents detailing their use of early entry policies. Such practices were presumably adopted to attract prospective students to the school (ACME, 2011).

Other anecdotal evidence suggested that some schools used early entry as a means of reducing the pressure on students by spreading GCSE assessments over multiple examination series, a practice that some students were reportedly in favour of (e.g. see Curtis, 2009). Indeed, Putwain (2008) hypothesised that in addition to spreading the assessment burden, early entry could reduce the stress experienced by students as it provided examination practice. The extent to which the former approach would be effective is clearly linked to whether early entry was coupled with multiple entry though.

In addition to using early entry as a standalone approach for the most able students (as discussed earlier in this chapter), another slightly different motive for schools to use early entry included situations when there were concerns about whether a student would

remain in education until the end of compulsory schooling. In such instances, early entry might have been in the best interests of the student, potentially allowing them to gain a qualification that they otherwise would not (ACME, 2011; Ofsted, 2013; Pope and Noyes, 2011). Such situations were likely to be relatively rare, however, and were not likely to have contributed significantly to the observed increase in early entry, particularly since there is nothing to suggest that this was a new approach.

The anecdotal evidence on why schools entered their students early and on multiple occasions was later supplemented by a report published by Ofsted (2013) that aimed to explore why schools were using early entry, which schools were using early entry, and the possible effects on attainment. Using a survey methodology and based on responses from 113 schools, Ofsted (2013) reported that the desire to raise standards and improve student motivation were the key drivers of early entry, with other influential factors being the desire to allow students to focus on other subjects or take additional subjects. Schools considered the key benefit of early entry to be its motivational effect (for students), but disadvantages included students needing to re-take assessments and disappointment with results (some schools also reported that early entry was demotivating, offering a contrasting view to those schools that saw it as motivational).

While this report canvassed the views of a relatively large sample of schools (113 schools) and provides some insight into the drivers of early entry (which was previously lacking), there are several perceived limitations that suggest that the evidence should be interpreted with caution. First, it was notable that the survey only focused on early entry and there was no specific discussion of multiple entry (except with reference to students re-taking). This was perhaps because early entry was seen as key at the time the survey was developed (the survey was conducted in early 2012), yet it means that the wider context within which early entry was operating may not have been fully considered (see Chapter 1). Second, there are concerns around the methodology. The study used a survey with closed questions to collect data from respondents and it was apparent in the

final report that the proportion of 'other' responses to certain questions was high, particularly questions relating to which students entered early and the factors that influenced the decision to enter early – questions that are of key interest here. Furthermore, details of these 'other' responses were not reported, so it is not possible to know what they consisted of. Clearly though, additional factors were guiding schools' approaches than those available to respondents in the questionnaire, suggesting that the questions or the available responses were poorly constructed. Finally, it could also be questioned how candid schools would have been in their responses since the report was administered by Ofsted who are also responsible for inspecting schools (see Chapter 1). While this is not intended as a criticism of the work, it is worth bearing this in mind when interpreting the findings. Indeed, the absence of any discussions relating to performance and accountability measures in the report could be influenced by this.

Summary of possible factors contributing to EMEE

To summarise, the increase in EMEE was considered to be influenced by a number of policy decisions, primarily the increased focus on performance measures for accountability purposes. Despite this, it is also likely that there were different drivers within individual schools and for different students, depending on their exact circumstances. While some schools may have used early entry to strategically support certain students, others appeared to be using it for a wider range of students (Ofsted, 2013). This suggests that the drivers of EMEE were likely to be complex and multi-faceted.

It is also worth noting that around the time that EMEE was increasing there was little guidance or policy relating to EMEE and any that was published seemed to have little effect on uptake. For example, during the reform of GCSE mathematics qualifications for first teaching in 2010, QCA (2009) published a document titled 'Changes to GCSE mathematics' that appeared to discourage wholescale early entry. The document stated that,

> *"Candidates who take GCSE early and achieve a lower grade than A* are less likely to continue their study of mathematics post-16 than students who achieve their full potential in mathematics at age 16. In other words, for candidates who may achieve lower grades through early entry, it would be better to delay entry and give them a richer experience of mathematics and the opportunity to achieve a higher grade"*

Nonetheless, the increase in EMEE continued. A similar piece of guidance was published on the DfE website in 2012, indicating that early entry should only be used in certain circumstances, yet again EMEE continued to rise.

> *"Exams taken at the end of the course will not necessarily be taken at the end of Year 11/Key Stage 4. If appropriate, pupils can still complete full GCSE courses before then. However, there has been a significant increase in the number of pupils being entered early for GCSE exams in recent years, and statistics show that this may not help pupils achieve their full potential. Teachers should be sure that early completion of GCSEs is right for individual pupils"*

Given the lack of any strong guidance on early entry it is perhaps not surprising that in response to policy changes early entry and multiple entry became more common. Nonetheless, concerns grew as the prevalence of EMEE increased. These concerns were generally related to the possible effects of EMEE for various stakeholders in the assessment system, which are explored in the following section.

What are the possible effects of EMEE?

At the time that this thesis was planned there was little empirical evidence exploring the possible effects of EMEE, despite much speculation in the subject communities and the media regarding what were mainly perceived to be negative consequences (e.g. see ACME, 2011;The Mathematical Association, 2010). This changed in subsequent years as a number of reports were published that focused on the potential effects of EMEE for students, in particular on their attainment.

While the effects of EMEE were arguably most significant for the students themselves they were also likely to extend beyond this, with implications for other key stakeholder groups such as teachers, schools and awarding bodies. Such effects were likely to have been both positive and negative, and the following section outlines the possible consequences of EMEE for each stakeholder group, drawing on the evidence that is available. Since much of the literature relating to EMEE concerns GCSE mathematics it is this subject that is the main focus. This is also in keeping with the overall focus of this thesis (see Chapter 1).

The effects on students

Attainment

Perhaps the most significant consequence of EMEE is the effect that it may have on students' attainment, as this is likely to influence future progression to further education and employment.[36] A number of sources argued that early entry would have a detrimental effect on attainment, with critics suggesting that early entry could result in students achieving a lower grade than they would otherwise be capable of (ACME, 2011; DfE, 2011; Noyes, Drake, Wake and Murphy, 2010; Winter, 2001). In turn, this could affect a students' ability to gain access to post-16 and University level education, where certain courses and Universities require the highest grades at GCSE as well as at A level (Cookson, 2012; Henry, 2008; Paton, 2010).

Much of the evidence considering the implications of EMEE on attainment focused on either early entry or multiple entry. While this might suggest that more holistic evidence relating to the effects of EMEE is lacking, this lack of evidence is perhaps not surprising when the complexities of generating such evidence are considered. Within the definition of

[36] It is worth noting that we are discussing attainment here in terms of the grade that is achieved in the GCSE assessments. The issue of how EMEE might interact with students' learning and performance in later assessments is discussed later in this chapter.

EMEE used in this thesis (see Chapter 1), a number of different routes through GCSE assessments are possible: students could enter early at a number of different time points (e.g. throughout Year 10 or mid-way through Year 11), and they could re-sit multiple times (again at different times). Thus, making any comparisons between students that entered early and on multiple occasions and those that did not is unlikely to be sufficient to understand the effects of EMEE due to the nuances involved. As such, the discussion here generally focuses on the evidence relating to early entry and multiple entry separately. Nonetheless, when considering the evidence relating to early entry it is important to remember that this might interact with multiple entry, depending on the route that students took through the assessments.

A detrimental effect of early entry for the most able?

The evidence relating to the potential effect of early entry on attainment has grown as the prevalence of early entry rose. One of the earliest reports was published by the DfE (2011) and related to analyses based on data from the June 2010 examination series. These analyses revealed that students who entered early were less likely than those entering at the end of Year 11 to achieve the highest grades (A* – B) in GCSE English and mathematics. While this analysis did not control for students' underlying ability (which may not have been comparable between the two groups), further analyses revealed that of the students with high prior attainment (levels 4 and 5 at KS2), the proportion achieving the highest grades at GCSE (A* – B) was greater amongst those that did not enter early. This suggests that early entry might have had a detrimental effect on attainment for students with high prior attainment at KS2.

The report also considered the effect of early entry for lower attaining students and here the relationship appeared to be different. Of those students not achieving a level 4 at KS2, a slightly higher percentage of early entry students achieved a grade C as their final grade compared to non-early entry students (DfE, 2011). However, the majority of these students still achieved a grade D or lower. This suggests that in terms of achieving a

grade C or above, any benefits of early entry for students with lower prior attainment are negligible.

The potential for early entry to have a detrimental effect on outcomes for students with high prior attainment was also observed by Ofsted (2013), who reported similar findings to the DfE (2011) based on data from the following academic year (see also Gill, 2013; Ofsted, 2012). When considering students with high prior attainment (level 5+ at KS2), Ofsted (2013) showed that in 2011 the percentage of students achieving A* or A was higher for students entering at the end of Year 11 in both English and mathematics compared to those entering early. A similar effect has been observed in other assessments: early entry to KS3 assessments resulted in fewer students progressing to the highest levels of attainment in English and mathematics, despite more students making the 'expected' level of progress (Node, Rutt, Schagen and West, 2007). The concern that students with high prior attainment would not achieve their potential if entering early was shared by anecdotal evidence from the teaching community. For example, a case study exploring learning trajectories in 14 – 19 mathematics revealed that one school rejected early entry as they believed it might risk achievement of the highest grades (Noyes and Sealey, 2011). While this view may not be shared by the whole of the mathematics community, it is unlikely to be an isolated case.

Although the evidence therefore suggests a detrimental effect of early entry on GCSE attainment for students with high prior attainment, it is also worth considering that there could be wider benefits of early entry. Thus, early entry could have benefits elsewhere in terms of a student's overall attainment across multiple subjects. Gill (2014) explored this through propensity score matching that compared the performance of early entry students in summer 2011 with matched counterparts on three measures: mean GCSE score, the chance of passing the five A* – C GCSE threshold, and total GCSE points score. The analyses showed that early entry had little benefit for individual students when considering mean GCSE score and comparing students that entered at least one GCSE early with

students that entered no GCSEs early. However, when considering total points score, students entering one GCSE early had a higher score than those not entering any GCSEs early (the opposite effect was observed for students entering more than one GCSE early though). Furthermore, when considering the threshold measure students that entered one GCSE early significantly out-performed those not entering any GCSEs early (by 2.8%).[37] Gill (2014) concluded that while early entry did not seem to benefit individual students, there seemed to be some success in getting students to achieve the threshold measure. He hypothesised that this might be because early entry was beneficial in getting students around the grade C boundary to improve their grade, but less beneficial in ensuring that higher achieving students reached their potential. Given the earlier discussions in this chapter, this seems plausible.

While the findings discussed above constitute the main published literature concerning the effects of early entry on attainment, there have also been efforts to compare the performance of students in different age groups by the awarding bodies, with a view to exploring the effects of maturation on performance for the purpose of maintaining standards in modular qualifications (see Heinrich and Pinot de Moira, 2012; Pinot de Moira, 2008). Although entering individual modules of a modular qualification might not constitute early entry as defined in this thesis (since modular qualifications encourage students to enter some modules at a younger age by their very nature), such analyses provide some insight into how students of different ages engage with assessments, which is relevant to the issue of early entry.

When comparing the performance of 15- and 16-year-old students on GCSE assessments, Pinot de Moira (2008) showed that performance at unit and subject level tended to differ across age groups. However, these differences were not consistent

[37] Note that for the low attainers and students from independent schools the opposite effect was observed though.

between subjects. In some subjects 16-year-olds out-performed their younger counterparts, while in others the opposite effect was observed. These differences could be a reflection of the content of each subject and the structure of the assessments within each qualification, as it is possible that differential performance by age is associated with certain content, questions or assessment types. Furthermore, they could be a reflection of the manner in which students were prepared for the assessments. Nonetheless, the findings suggest that there could be some effect of early entry on performance due to the age of the students.[38]

Improved performance through multiple entry or re-sitting?

There is less evidence considering the effect of multiple entry on attainment, but the effects are likely to be more complex. For example, students entering on multiple occasions are likely to have taken different routes through the assessments and certificated at different time points, meaning that they might be expected to perform differently on each attempt – for example, if levels of preparedness or motivation differ. It is therefore very difficult to consider directly how multiple entry might impact on performance. However, to help us here we can consider the literature relating to re-sitting, since multiple entry is considered to be a form of re-sitting (see Chapter 1).

One report considering the effects of multiple entry on outcomes was published by Vidal Rodeiro (2014), focusing on students that certificated twice in the same examination series. Although this research offers a much narrower definition of multiple entry than the one used in this thesis (see Chapter 1), it provides some insight into the potential effect of multiple entry on attainment when other factors should be relatively constant (e.g.

[38] Note that there is also a body of literature relating to season of birth effects and the possible impact on performance (see Sykes, Bell and Vidal Rodeiro, 2009, for a review). This literature is not discussed here since although there are some parallels with early entry (e.g. in terms of students of different ages sitting assessments), the implications of entering early extend beyond the age of the student (i.e. in terms of how they might be prepared for the assessments).

preparedness, teaching time in the subject, etc.). Vidal Rodeiro (2014) compared the attainment of students entering multiple times in the same examination series with single entry students.[39] The analyses suggested that when comparing the performance of single and multiple entrants with the same prior attainment, those with lower prior attainment who entered twice tended to perform better than those who entered once. However, for the students with higher prior attainment, the opposite effect was found.

To consider further the possible effect of multiple entry on attainment, Vidal Rodeiro (2014) used propensity score matching. This analysis showed that after matching (on a number of variables including gender, FSM, ethnicity, prior attainment, SEN, school type, language, and deprivation), multiple entrants marginally out-performed the single entrants in mathematics, but not in English. However, for both subjects the single entrants were more likely to achieve the top grades, while multiple entrants were more likely to achieve grades C or D. Overall, these findings suggest that there seems to be a small benefit of entering on multiple occasions for the lower ability students. It is possible that such effects are related to the effect of random measurement error in assessments that could increase the chance of students receiving a higher grade on one of their attempts (see Chapter 6 for a full discussion). Indeed, it is worth noting that of the students that entered two GCSE qualifications in the same examination series only half achieved the same grade on each qualification (although there are likely to be a number of factors that could influence performance when students enter twice in the same examination series such as differences in motivation; Vidal Rodeiro, 2014; see also Keeling, 2009).

Although there is limited evidence relating to the effects of multiple entry on attainment (and we have discussed the complexities of this), it is also worth drawing upon the

[39] The report also compared the grades that students achieved on multiple examination attempts in the same series. This is discussed in Chapter 6 where the possible impact of EMEE on national outcomes is considered.

literature relating to re-sitting since there are clear parallels (see Chapter 1). Much of the evidence concerning re-sitting in high stakes assessments in England relates to the re-sitting of individual modules or units within a qualification (e.g. see Vidal Rodeiro and Nádas, 2012). Here, the impetus to re-sit is presumably to improve performance on the module and therefore the final grade that a student achieves. While for some students the impetus to certificate on multiple occasions at GCSE might be the same, other students might re-sit due to a blanket re-sitting policy in the school or to ensure that they keep studying a certain subject until the end of KS4. Thus, multiple entry is not necessarily directly comparable to re-sitting modules, yet some of the issues are likely to be the same. The literature relating to re-sitting reported here focuses specifically on curriculum related assessments from the English assessment context, though it is noted that there is other literature relating to the effects of re-sitting in other types of tests.

One of the key studies that considered the effect of re-sitting modular assessments on examination performance was published by Cambridge Assessment (Vidal Rodeiro and Nádas, 2012). Vidal Rodeiro and Nádas (2012) showed that for both GCSE English and mathematics students tended to perform better on their re-sit attempt, with between 30 – 50% of students in both GCSE English and mathematics improving their final unit grade. Other research has shown similar findings (Al-Bayatti and Jones, 2003; Spalding, 2009). Thus, it seems that the opportunity to re-sit modules can lead to improved outcomes. A similar effect could be observed in multiple entry if students were better prepared for the later entry attempts and sufficiently motivated and engaged.

The potential to perform better on a second attempt could be due to a number of factors (see Vidal Rodeiro and Nádas, 2012, for a discussion). Students might not be well prepared on their first attempt, particularly if they have had a limited amount of time to prepare for the assessments (Vidal Rodeiro and Nádas, 2012). This could also apply in EMEE, particularly if students enter early in Year 10 after limited teaching time. Alternatively, students might lack the maturity to engage with the assessment and might

therefore perform below their potential on their first attempt, a criticism often levelled at modular assessments (Clarke, 1996; Taverner and Wright, 1997). As such, younger students might be at a disadvantage, resulting in them falling below the expected standard (Vidal Rodeiro and Nádas, 2012) – not because they are not able enough or prepared for the assessments, rather, that they have not developed the skills required of them to engage with and respond to the assessments as required. Again, this could apply in EMEE.

In addition, students might not be developmentally ready or able to engage with the subject content as required, which is likely to affect their learning. Indeed, this was observed in a study exploring the effects of early entry at KS3, with teachers reporting that some students were not developmentally ready to tackle the GCSE programme of study early (Node, Rutt, Schagen and West, 2007). The knowledge that a re-sit is available may also mean that students do not try their best on their first attempt (Vidal Rodeiro and Nádas, 2012). The same could be true in EMEE, since the stakes associated with each entry attempt are arguably lower.

Other possible explanations for students improving their performance is that the opportunity to re-take can lead to students learning more and therefore improving their performance, since they have the opportunity to receive formative feedback after each assessment. The same could be true in EMEE at certification level. Re-sitting might also allow students to progress by exposing them to the assessments at an earlier stage, thus providing examination practice and thereby increasing their confidence in a later sitting (see Putwain, 2008).

To summarise, the relationship between EMEE and attainment appears to be complex, and likely to depend on a number of factors that are influenced by the circumstances of individual students. While early entry might have a detrimental effect on attainment if teaching time is reduced or students lack the maturity to engage with the subject content or the assessments, early entry followed by multiple entry might be beneficial if it provides

students with examination practice or allows them to receive feedback that helps improve their performance on a later attempt. This means that drawing any overall conclusions based on the evidence available is difficult. Nonetheless, there is some evidence that early entry can have a detrimental effect for students with higher prior attainment, by preventing them from achieving their potential. However, there is also evidence that multiple entry in the same examination series can result in slightly higher outcomes overall for the lower ability students.

A further point to consider is that although the discussion here has focused on the effects of EMEE on attainment for individual students, there are likely to be wider effects at the overall cohort level. As discussed in Chapter 1, national outcomes for GCSE assessments are reported each August and are subject to intense media scrutiny. Thus, if early entry were to impact on the number of students achieving the top grades, or if multiple entry were to increase the number of students achieving the middle grades, then this might influence – and perhaps even distort – national outcomes. This is discussed fully in Chapter 6 so is not considered here, but it is clearly related to the potential for EMEE to impact upon the attainment of individual students.

Future progression and access to further education

In addition to influencing attainment at GCSE, EMEE might have knock-on effects for students beyond the grade that they achieve at GCSE, including on their future progression in a subject and their access to further and higher education (see Cookson, 2012; Paton, 2010). Since much of the evidence concerning future progression relates to early entry, that is the focus in this section.

The key concern is that any under-achievement through early entry could have implications for progression onto A level courses. For mathematics, it is argued that higher attaining students who under-achieve through early entry could be put off studying mathematics related courses in post-16 education (ACME, 2011). This effect could be particularly acute for students who achieve below a grade A at GCSE, given that students

are more likely to continue to A level mathematics if they have achieved a grade A or A* (QCA, 2009). This premise is supported by evidence showing that while almost 75% of students who achieved an A* at GCSE continued to study mathematics in post-16 education, this figure dropped to 35% of those who achieved a grade A and 5% of those who achieved a grade B (DfE, 2011). Hence, higher ability students who enter early but achieve lower grades might not continue studying mathematics or mathematics related courses in the future, potentially to their detriment and to the detriment of the wider society (The National Centre for Excellence in the Teaching of Mathematics, 2011), particularly given the well documented shortage of mathematics specialists in the UK (Noyes, 2009; Noyes, Wake and Drake, 2011). This could particularly be the case in schools where the entry policies onto A level mathematics courses requires at least a grade B at GCSE (see Noyes and Sealey, 2011).

Despite such concerns, there is some evidence to suggest that there is no association between early entry and progression to A level in English and mathematics. Gill (2013) explored future progression and future A level performance in English and mathematics depending on when students sat their GCSE, using data for students that completed GCSEs in summer 2009 and 2011 (see later discussion in this chapter about the possible effects of EMEE on future performance). The analyses suggested that there was no association between when students entered and the likelihood of continuing to A level, when considering both the timing of a student's first entry attempt and their final entry attempt. Indeed, there was some evidence that students who entered early were more likely to progress to A level. However, as outlined by Gill (2013), these analyses must be treated with caution since they consider progression based on the grade that each student achieved at GCSE, and not each student's *actual* ability. Thus, if a student entered early and achieved a lower grade than they were capable of, they may still have had a higher chance of progressing to further study than a student who entered later and achieved the same grade (but were actually of lower ability). It is also worth considering that these

analyses cannot consider the possible effects of multiple entry that may influence student's engagement with a subject (see later discussion). Thus, this study, the only one of its kind that was identified, can only provide limited insight into the implications of EMEE for future progression.

Learning and future performance

Another related issue is a possible effect of EMEE on students' learning and future performance, though this is largely speculative since there is little evidence to draw upon (the focus here is again on early entry). While there should be no doubt about the importance of the grade that a student achieves at GCSE, it is important to note that this grade is only a snapshot of what they can achieve at a particular time. Hence, it is possible that early entry could result in positive outcomes in terms of students' learning and future performance, even if they achieved a lower grade through early entry. For example, students who follow an enriched curriculum through early entry and study additional material might have learnt more by the end of Year 11 than their peers who did not enter early. As such, a lower grade might not necessarily be a reflection of what a student knows and can do by the end of Year 11, despite it having negative connotations.

Gill (2013) considered the possible effect of early entry on future performance at A level using data for students that completed their GCSEs in summer 2009 and 2011. Focusing on English and mathematics, the analyses suggested that students entering early did not appear to be disadvantaged in terms of their A level performance. However, Gill (2013) noted that information on what students did following their early entry attempt was lacking so could not be accounted for in the analyses. For example, some students might have studied an enriched curriculum which may have resulted in improved performance, while others might have had a gap in their learning (see later discussion in this chapter). Consequently, any conclusions are limited. Furthermore, these findings only really relate to early entry, so the effects of EMEE are less clear.

Although only speculative, it is possible that early entry could have other implications for students' learning in addition to their future performance in a particular subject. For example, early entry for particular subjects might be beneficial as it could allow students to focus upon one subject in greater depth for a certain period of time before moving on to study other subjects. Indeed, in a report by Ofsted (2013) one of the key perceived benefits of early entry by schools was that it allowed students time to study additional subjects.

Early entry might also impact positively on students' later learning by allowing them to apply the knowledge they have gained through early entry when studying other subjects. Conversely, however, concentrating on one subject could be detrimental as it may result in less synoptic learning and less transfer of knowledge between subjects. These effects are likely to depend upon which subjects students study when, and the subject content itself, so would be difficult to quantify.

The effects on a student's learning will also depend upon the provision that schools offer their students who enter early and whether early entry is coupled with multiple entry. Evidence suggests that some schools had poorly thought out plans for students that entered early and did not re-sit, or offered no provision at all (Pope and Noyes, 2011), meaning that students may have experienced a gap in learning during Year 11 that is likely to have influenced their intentions to study certain subjects in post-compulsory education and their subsequent performance (ACME, 2011; DfE, 2011; Noyes, Drake, Wake and Murphy, 2010). In other instances, schools reported well laid plans for supporting students who entered early including enrichment opportunities, acceleration to A level, and the study of additional qualifications (see Bond, Green and Jaworski, 2009; Noyes and Sealey, 2011).

Within individual schools the provision for students is also likely to differ depending upon a students' attainment at GCSE, in terms of the actual grade that they achieved. A survey of 368 mathematics departments in summer 2010 revealed that while the high attaining

students typically had access to additional mathematics at a higher level following early entry, this provision declined in line with students' attainment at GCSE, with the majority of grade D and E students studying the same mathematics or more mathematics at a similar level following early entry (Noyes, Drake, Wake and Murphy, 2010). Perhaps of more concern, however, were the numbers of students that had no access to mathematics following early entry: of the students who achieved a grade A through early entry almost 10% studied no further mathematics, a figure that rose to just over 23% for grade C students (Noyes, Drake, Wake and Murphy, 2010).

The potential for students to experience a gap in learning following early entry could be particularly detrimental for future performance if it was coupled with students not studying the whole GCSE programme of study due to early entry (DfE, 2011). Several subject commentators argued that schools might be tempted to prioritise certain aspects of the curriculum in order to ensure that students secured a grade C when entering early (ACME, 2011; IMA and LMA, 2011). While this is likely to have a detrimental effect on students who study mathematics related courses in post-16 education, more importantly, it is also likely to result in schools not providing students with the full education to which they are entitled (ACME, 2011). This effect could be compounded by teachers focusing on passing the test rather than ensuring that students understand the mathematics that they are being taught (DfE, 2011). For example, it is argued that students who enter mathematics early might develop a 'right answer' attitude, thus face difficulties when required to investigate or question (Winter, 2001).

Motivation and engagement

While it is clear that EMEE could have an effect on students' attainment, learning and future progression, it was also claimed that EMEE could influence less tangible factors such as a students' motivation and attitude. Such effects are likely to be associated with the propensity for students who entered early to also enter on multiple occasions, particularly those who were predicted to achieve a grade C or D (Noyes, Drake, Wake and

Murphy, 2010). For some students there may have been benefits of multiple entry beyond any related to the implications for attainment (see earlier discussions relating to the effects of re-sitting on attainment). Indeed, for students who lack self-esteem EMEE might have given them the motivation and confidence to improve their original grade in a subsequent examination series (Association of School and College Leaders [ASCL], 2011). Furthermore, EMEE might have allowed borderline students the opportunity to enter once at foundation tier then engage with more difficult material and re-sit the higher tier in a later examination series (e.g. see Curtis, 2009). However, there are also likely to be negative consequences. It is argued that repeated entry to GCSE mathematics could negatively affect students' perceptions (Noyes, Drake, Wake and Murphy, 2010), causing them to be disinclined to study mathematics related subjects in post-16 education (The Mathematical Association, 2010).

Opportunities

Although EMEE may result in negative effects on students' attitudes and motivation, perhaps of more concern are instances involving students who under-achieve in their first attempt but are not given the opportunity to re-sit and improve their final grade. The opportunities for students to re-take following an early entry were shown to vary quite considerably between schools within a sample of 692 secondary schools in England in 2011: 35% of respondents reported that students who entered early and achieved a grade C would 'sometimes' get a chance to re-take, with a further 6% reporting that students would not be allowed to re-take (Centre for Analysis of Youth Transitions, 2011).

The opportunity for students to re-take is also likely to be related to students' attainment in their early attempt at GCSE: evidence suggests that it is those who are close to the C/D borderline who are more likely to re-take following early entry (Noyes, Drake, Wake and Murphy, 2010), raising concerns about the higher attaining students who did not achieve their potential on an early attempt (see earlier discussion in this chapter). As well as students not being given the opportunity to re-sit, there are also likely to be those who

were not motivated to improve their final grade, despite being able or having the opportunity to do so (Pope and Noyes, 2011). Both these situations are likely to result in under-achievement that could negatively influence future progression and students' attitudes and motivation.

Positive effects of standalone early entry?

Despite the perceived negative consequences of EMEE there were also likely to be benefits in certain situations. As discussed previously in this chapter, standalone early entry can be beneficial when used to stretch the most able students, by allowing greater breadth of study and curriculum enrichment. Furthermore, standalone early entry may be beneficial when there is a risk of a student leaving or being excluded before the end of compulsory education (ACME, 2011; Pope and Noyes, 2011), since it enables students to gain a qualification that otherwise they might not have. These instances are likely to be relatively rare though and have different drivers and implications to the widespread use of EMEE for a range of students.

Summary of possible effects of EMEE for students

To summarise, the possible effects of EMEE for students are likely to be complex and dependent upon an individual students exact circumstances and the opportunities available to them. While a body of literature exists considering the effects of early or multiple entry on attainment, there are likely to be other effects that are less well researched – for example, the effect on learning and future progression. Gathering such evidence is likely to be complicated by the sheer number of routes that were available through GCSE mathematics assessments when EMEE was on the rise (see Chapter 1): students were able to enter early in multiple examination series each academic year, and were also able to re-sit multiple times on different occasions.

It is also worth considering that although the discussions relating to the possible implications of EMEE have largely been negative, the importance of the grades that

students achieve cannot be underestimated, particularly with regards to accessing further University education and employment. Indeed, many Universities and employers cite grade C in GCSE English and mathematics as their minimum entry criteria and this grade is generally described as a 'good' pass (see DfE, 2010). Thus, although there are possible negative effects of EMEE (such as the effects on motivation and engagement), for some students these might be off-set (in the short-term at least) by the ability to enter on multiple occasions if this facilitates achievement of a grade C, particularly in a key subject such as mathematics.

The effects for schools

In addition to the effects of EMEE on individual students there were also likely to be wider issues for schools and teachers, including those related to logistics and resources. One issue that has featured already in this review is the need to provide suitable provision for students who completed their GCSE early and for students who intended to re-sit after entering early. Both these situations were likely to have implications for timetabling arrangements in schools. In the former case, some students may go straight on to study at A level following early certification, which has the potential to cause difficulties if the A level timetable does not align with other commitments at GCSE (Noden, Rutt, Schagen and West, 2007). Equally, re-sitting the GCSE or studying an additional enrichment qualification might interfere with the scheduled timetabling of other GCSE subjects. In a study exploring mathematics pathways in Year 11, one school reported discontinuing EMEE as students had difficulties re-taking GCSEs and studying a free-standing mathematics qualifications (FSMQ) alongside this (Bond, Green and Jaworski, 2009). Furthermore, in a survey of 368 mathematics departments a number of teachers rejected early entry as it resulted in an increased workload for teachers and timetabling restrictions (Noyes, Drake, Wake and Murphy, 2010). The logistical issues associated with EMEE are likely to be intensified when students transfer between teaching groups or schools and are

at different levels to the students that they are joining. This might result in students being ahead of their peers and thus repeating material, or indeed trying to catch-up or being unable to transfer between ability sets (Noden, Rutt, Schagen and West, 2007).

Alongside the logistical problems associated with EMEE there were also likely to be resource implications for schools given the associated time and monetary costs (DfE, 2011). ACME (2011) believed that increased expenditure on examination fees resulting from EMEE was a poor use of resources, particularly considering the rise in expenditure by schools on examination fees each year that was highlighted around the same time (Ofqual, 2011). These concerns were raised at a time when the education budget was under pressure due to Government cuts (Chowdry and Sibieta, 2011),[40] with many schools reporting decreasing budgets and assistance from local authorities (e.g. Richardson, 2011). While EMEE might therefore increase the squeeze on resources, for some schools the extra costs might have been worthwhile if it helped the school avoid falling below floor standards and the associated consequences that this brings (see Chapter 1).

The effects for awarding bodies

While the effects of EMEE are likely to be most significant for students and schools, there are also potential implications for awarding bodies, the main issue being related to standards. The process of maintaining GCSE examination standards is driven by prior attainment based predictions, as outlined in Chapter 1. This approach is most reliable when the cohort of students is large and remains 'stable' over time, meaning that the same type of students are sitting the qualification in each examination series. Hence, there are obvious issues in the presence of EMEE since changes in entries could

[40] Note that the situation has not changed since.

compromise the stability of the population and therefore the reliability of any statistical indicators. The implications of EMEE for examination standards are considered in detail in Chapter 8 so are not discussed further here.

In addition to the effects on standard setting there were also likely to be financial implications for awarding bodies. Increased EMEE is likely to result in increased income from examination fees for awarding bodies (see DfE, 2013b), a topic that received attention in the media due to the increasing expenditure by schools on examination fees at the time EMEE was increasing (see Paton, 2012). Thus, there may have been perceptions that awarding bodies supported EMEE as a means of generating extra income, despite concerns about these practices being voiced by such bodies (e.g. Hall, 2012). Such perceptions could impact negatively on public perceptions of the awarding bodies and the wider assessment system though.

Early and multiple entry in other jurisdictions

This literature review has focused on EMEE in the English assessment context, yet it is also worth considering evidence from other jurisdictions. Of most relevance is the recent situation in Wales, where the prevalence of EMEE to GCSE assessments rose in prominence between 2012 and 2017 (Qualifications Wales, 2017). This prompted Qualifications Wales (the examinations regulator for Wales) to initiate a research study exploring the possible drivers of EMEE within schools, the report of which was published in October 2017 (Qualifications Wales, 2017). Although this report was published as this thesis was nearing completion and therefore did not influence the work presented here, there are obvious parallels between the situations in England and Wales (see Chapter 1).

The report by Qualifications Wales (2017) discussed the findings of a qualitative study (similar to the study reported in Chapter 5 of this thesis) that considered the factors influencing schools' approaches to EMEE, drawing on interviews with 59 professionals (e.g. teachers, heads of English, Welsh, or mathematics departments, head teachers,

exams officers, and regional consortia staff) in 21 schools and four regional consortia in Wales. The findings suggested that the two key drivers of EMEE were the desire to raise student attainment and to improve schools' results in relation to performance measures, the latter reflecting some of the anecdotal concerns discussed earlier in this chapter. However, the picture was also quite complex: some schools used standalone early entry to stretch the more able students, while others used early entry for disadvantaged students to increase the chance of them achieving a qualification, with respondents perceiving there to be benefits of these approaches. Thus, EMEE served different purposes in different situations.

The report also discussed respondents' perceptions of the positive and negative effects of EMEE, which again reflected some of the issues already discussed in this chapter. Some respondents thought that EMEE was beneficial since it increased student motivation, facilitated providing formative feedback to students (external tests were seen as preferable to internal tests), allowed more targeted teaching, increased aspirations, and spread the burden of assessment over multiple examination series. However, other respondents thought that EMEE could lead to disengagement and a loss of confidence if students did not perform as expected, and cited negative implications for teacher workload, the burden of assessment, and school expenditure on examination fees. Furthermore, it was thought that EMEE could reduce teaching time and negatively impact on students' future progression: early entry might result in students not achieving their potential, while multiple entry might be negatively perceived by some of the more selective Universities.

The report concluded that although the priority for respondents was to ensure the best for their students and schools, different approaches to entering students to assessments were used in differing circumstances. Thus, there was not a single approach that would suit all students and schools and any policy changes should be carefully considered in light of this.

Ultimately, when the report was published changes were made to accountability measures such that from summer 2019 only a student's first entry attempt would count in school performance measures (Welsh Government, 2017), mirroring the approach adopted in England since September 2013 (DfE, 2013a). Although the timescales for these changes were more generous than those in England (where the changes were effectively made overnight), the changes were not met with universal support publically and concerns were raised about the implications for certain students (see Wightwick, 2017). The implications of such changes for EMEE in Wales was not known at the time of writing, but the evidence from England suggests that EMEE will decline fairly rapidly once the changes come into effect.

In contrast to the UK context, other assessment systems (such as the US) use early entry in a different way, as here it is more common for students to be accelerated a whole academic year. As such, students are not entered early in the same way, since they have studied alongside their older peers for a sustained period of time. Thus, any evidence relating to early entry in the US is likely to be less relevant so is not considered here.

Summary and perceived limitations of existing literature

To summarise, as the ideas for this thesis were developing the literature relating to EMEE was limited, suggesting that a more detailed investigation of EMEE was required. However, as this thesis progressed the literature has expanded to consider the possible effects of early entry on attainment, progression and future performance, and the possible implications of multiple entry in the same examination series. While this has led to a wider body of literature considering EMEE, much of the evidence still resides in the grey literature and there remains a number of gaps that this thesis can contribute to.

The first relates to the trends in EMEE. As outlined in this chapter a number of studies have considered trends in early entry, but it is notable that many of these studies have focused on a relatively short time period. As such, while the number of early entries over a

longer time period has been considered (e.g. see DfE, 2011), there is less evidence relating to the trends over time at the student and school level. Furthermore, since the different analyses that have been conducted have typically used slightly different datasets or definitions, it is not always possible to make any direct comparisons between studies relating to different time periods. This thesis therefore considers trends in early entry by undertaking a more detailed analysis over a longer time period than has previously been available. This includes considering trends in the type of students that entered early and how they performed, as well as other factors such as their tier of entry – evidence that has not previously been available. The chosen time period for these analyses (covering five years between 2007 and 2011) represents the point that these analyses were conducted and the point at which early entry was rising.

A second gap in the literature relates to the possible drivers of EMEE. Much of the evidence relating to EMEE has used quantitative methods to consider trends in entries or the effects on attainment, but less is known about the factors influencing the rise in EMEE, other than that based on anecdote. This was the case as the ideas for this thesis were developing, and remains so now. Furthermore, the limitations of the only real attempt to gather such evidence, a report published by Ofsted (2013), have already been discussed. The perceived limitations of existing evidence suggests the need for alternative methods to explore the drivers of EMEE in greater depth. As such, a qualitative approach was used in this research (see Chapter 3 for further discussion of the methodology).

Finally, while there is literature considering the implications of EMEE for individual students, there are other possible consequences that are yet to be explored, including implications at the wider system level. This interacts with the implications for individual students (indeed any implications for students are likely to feed in to implications at the system level), yet there are other considerations as well. This thesis therefore explores the implications of EMEE for examination outcomes and standards. The former includes a consideration of the trends in multiple entry, which provides the necessary context for

considering the implications for outcomes, and the latter includes a consideration of the implications of EMEE for how examination standards are defined and maintained. Thus, the literature relating to definitions of examination standards provides the theoretical context for this thesis.

Research questions

Given the existing literature and the perceived limitations outlined above, this thesis addressed five key research questions (RQs):

> RQ1) What are the key trends in early entry to GCSE mathematics assessments between 2007 and 2011?
>
> RQ2) What influences schools' entry approaches to GCSE mathematics assessments, in particular, to entering students early and on multiple occasions?
>
> RQ3) What characteristics at the school and student level are associated with multiple entry to GCSE mathematics assessments?
>
> RQ4) What are the implications of EMEE for examination outcomes in GCSE mathematics assessments?
>
> RQ5) What are the implications of EMEE for maintaining examination standards?

While these are the broad over-arching research questions that underpin the work in this thesis, there are also a number of subsidiary research questions that are provided in the appropriate chapters.

Chapter 3 – Research design: using mixed methods

Overview

This chapter provides an overview of the research methodology and philosophical assumptions underpinning this thesis, including a rationale and justification for the approach taken. The research explores EMEE to GCSE mathematics assessments and the implications for maintaining examination standards. Specifically, it explores trends in EMEE, the drivers that influence EMEE in schools, and the consequences of EMEE for examination outcomes. The findings are considered in light of the theoretical literature relating to how examination standards are defined.

The research adopts a multiphase mixed methods design containing two quantitative phases and a qualitative phase. A description of the methodology is provided, alongside a consideration of the relative strengths and weaknesses of this approach. Subsequent chapters contain further details of the methodology in each strand of the research, including a consideration of data collection, sampling, and analysis techniques (see Chapters 4 – 7). As such, this chapter focuses on the overall research strategy, the use of mixed methods, and the quantitative and qualitative methods that were chosen.

The chapter also includes an overview of the ethical issues in this research and a reflexive account of the role and influences of the researcher. Such issues are particularly apparent in relation to the qualitative phase of the research, but are also considered to underlie the overall rationale and approach to the research. Primarily, these relate to the position of the researcher as an awarding body employee as the work in this thesis began, and the sponsorship of the research by an awarding body and the qualifications regulator.

Research design and methodology

The use of a mixed methods approach

This research combines quantitative and qualitative methods and is thus considered a mixed methods design. Mixed methods research is considered to be a more recent paradigm that has developed due to the appeal that combining quantitative and qualitative techniques offers (Teddlie and Sammons, 2010). Within mixed methods research a large number of research designs are possible, often making mixed methods approaches challenging for researchers (Creswell and Plano Clark, 2011). Indeed, research design is typically considered one of the unresolved challenges in mixed methods research (Teddlie and Tashakkori, 2003). Despite this, three key dimensions of mixed methods research are typically highlighted (e.g. Creswell, 2011; Leech and Onwuegbuzie, 2009): the level of mixing across methods (partial or fully mixed), the sequencing of the research (sequential or concurrent), and the emphasis on the two methods (equal or dominant).

A large number of mixed methods research designs contain two phases and several typologies have been developed to describe these approaches (e.g. Creswell and Plano Clark, 2011; Leech and Onwuegbuzie, 2009). However, the present study contained three strands in a time sequence – a quantitative strand, followed by a qualitative strand, then a final quantitative strand (see Figure 4). As such, the first phase entailed a quantitative analysis of national data to evaluate trends in early entry (to address RQ1); the second phase a qualitative study with teachers in schools to explore their rationale for EMEE (to address RQ2); and the third phase a quantitative analysis of national examination data and simulations to consider the implications of EMEE for examination outcomes (to address RQ3 and RQ4). The findings from these studies are then considered in light of the theoretical literature relating to how examination standards are defined (to address RQ5). This approach to mixed methods research has been described as a multiphase design (Creswell and Plano Clark, 2011), where the research contains a series of strands

or separate studies that address a common research topic (Creswell, 2011). Research designs that alternate between methods, as in the present case, have also been described as a ‘sandwich’ design (Sandelowski, 2003).

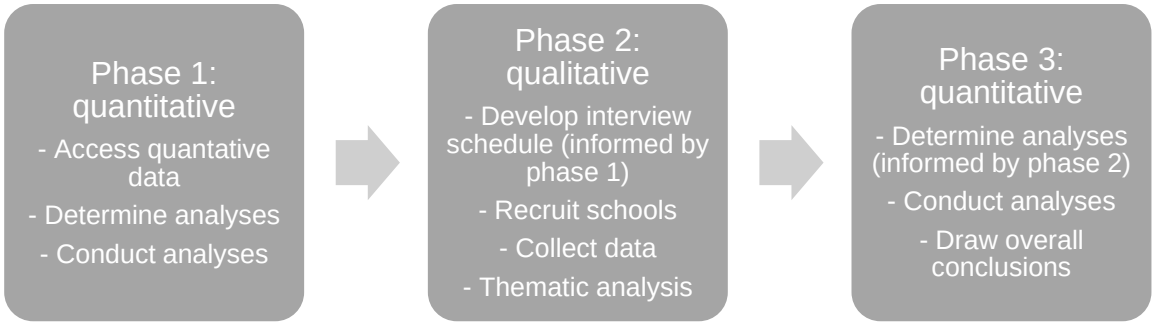


Figure 4. *Overview of research design*

In terms of the dimensions of mixed methods approaches, this research iterates between methods within separate stages of the research, places equal emphasis on both methods, and proceeds in a sequential manner. This allows each stage of the research to address separate research questions, yet also allows each stage of the research to inform subsequent stages (Creswell and Plano Clark, 2011), an approach that is facilitated by the multiphase design. Mixed methods and multiphase designs are considered complex (Creswell, 2011) and there are a number of issues that require consideration. Thus, the benefits and challenges of both mixed methods designs, and multiphase designs, are discussed later in this chapter.

Philosophical assumptions

This research adopts a pragmatic approach. This is a common stance in mixed methods research (Tashakkori and Teddlie, 2003) that avoids researchers being limited to certain methods or approaches (Feilzer, 2010; Johnson and Onwuegbuzie, 2004). Pragmatism therefore avoids the debate that divides positivists and constructivists (Hanson, 2008) and does not dictate the methods that should be used. Rather, it promotes the use of the most appropriate approach for what the research intends to find out (Feilzer, 2010; Johnson

and Onwuegbuzie, 2004). As such, the focus is on the problem that is being researched (Feilzer, 2010) and whether the research has contributed to the knowledge that the researcher seeks (Hanson, 2008).

Adopting a pragmatic approach facilitated the use of multiple approaches to address the research questions outlined in Chapter 2. Quantitative techniques were therefore used to explore trends in EMEE, the student and school characteristics associated with EMEE, and the implications for national outcomes, while qualitative methods were used to gain a deeper understanding of schools' approaches to entering students to assessments. This approach brings several benefits to the research, as outlined below.

The benefits of a mixed methods approach

The use of a mixed methods approach in this research was driven by many factors. Primarily, a mixed methods approach was chosen since it aims to "*provide a better understanding of the research problem and question than either method by itself*" (Creswell, 2011). Thus, mixed methods can expand the research beyond what can be achieved from using a single method (Bryman, 2006), by utilising the strengths of both methods. In the present case, quantitative methods (in both the first and third phase) allowed the analysis of large scale assessment data and outcomes. This entailed the use of secondary datasets – all awarding body national examination data and the National Pupil Database (NPD) – the merits and challenges of which are discussed later in this chapter. Qualitative methods complemented the use of large scale datasets, by providing a more in-depth and *"complex picture of the situation"* (Creswell, 2011). Although the qualitative strand was informed by the quantitative findings in the first phase, the qualitative phase was still largely exploratory, since the existing evidence concerning EMEE entry is limited or anecdotal (see Chapter 2). This supports the use of qualitative methods, since they are typically favoured as "*...the best strategy for discovery, exploring a new area…*" (Miles and Huberman, 1994). Furthermore, qualitative methods offer a

flexible approach to research, allowing the methods and procedures to be adapted throughout the duration of the study (Miles and Huberman, 1994).

Using both quantitative and qualitative methods allowed different research questions to be addressed (Bryman, 2006; Teddlie and Sammons, 2010), while also allowing each stage of the research to inform latter stages. For example, the initial quantitative phase showed trends in EMEE that were considered in the subsequent qualitative phase, while the qualitative phase allowed a more informed approach to the second quantitative phase. Thus, the sequential nature of the research allowed each phase to complement and inform the data collected in the other phases, and allowed subsequent stages to build upon the findings of the previous phase. Given the exploratory nature of aspects of the research this was considered beneficial, since it allowed subsequent stages of the research to be adapted, depending upon the findings of the previous phase.

The use of a multiphase design is considered to offer further benefits to those gained from a two-phase mixed methods approach. Primarily, this relates to the opportunity to conduct multiple research projects within an overall programme of research (Creswell, 2011). For example, a multiphase design offers a flexible approach to research that contains a series of interconnected research questions; allows researchers to complete and report on individual studies within the overall strand of research while other studies are progressing; and provides an overall framework in which to conduct multiple studies addressing a common research theme (Creswell and Plano Clark, 2011).

The challenges of a mixed method approach

Despite the benefits outlined above, mixed methods research and multiphase designs are challenging (Creswell and Plano Clark, 2011). First, the researcher is required to use and understand both quantitative and qualitative methods (Creswell, 2011). While in the present case the researcher had prior experience of using both methods in the field of interest, my expertise was weighted towards the quantitative elements, thus generating

additional challenges for the qualitative strand. A second issue relates to the integration of data (Creswell, 2011), since the quantitative and qualitative studies can generate distinct and opposing findings. Thus, although quantitative and qualitative methods are used in separate stages of this research, the overall findings must be integrated into a coherent whole.

Alongside the challenges of using mixed methods, a multiphase design can offer additional complexities, since the researcher must also be concerned with ensuring an overall coherence to the research. Thus, separate studies must combine meaningfully to address a common research theme (Creswell and Plano Clark, 2011). The researcher must therefore be concerned with ensuring that the research questions within each phase contribute to the overall research aim, and that each set of research questions builds on the findings from previous phases (Creswell and Plano Clark, 2011). This can involve adapting prior research designs and questions as the study progresses, a potential strength and challenge of this approach. While this can allow latter stages of the research to focus on particular issues that are identified in earlier stages of the research (leading to a better informed programme of research), it can also mean that research plans must be reviewed and adapted as the research progresses.

A final but important issue in relation to the multiphase element of this research concerns the duration of study, which has spanned over six years. During this time, there have been shifts in policy that have had significant implications for the behaviours being studied (see Chapter 1). For example, changes to accountability measures announced in September 2013 (DfE, 2013a) largely eradicated EMEE for large cohorts of students. As such, the research has progressed against a backdrop of policy changes, requiring the focus to adapt at times. This issue is a common reality for those researching in the field of education and educational assessment, due to the pace at which education reform typically occurs (Baird and Lee-Kelley, 2009). While this has been challenging it has also prompted the researcher to consider the findings of the studies within a wider context,

rather than just in relation to EMEE. Thus, while the research focuses on EMEE in the English assessment context, the findings extend beyond this (see Chapters 1 and 9).

The changing backdrop to the research has also had implications for the way in which the research has been presented (see Chapter 1). As such, rather than attempting to update earlier analyses with more recent data, the thesis is presented as a chronological narrative. This aims to ensure that the unfolding story of EMEE is not lost, and that the analyses provide an overview of the situation as it occurred in practice.

The use of secondary datasets

The two quantitative phases of this research (phases one and three) involved the secondary analysis of large scale numerical datasets. Secondary analysis can offer huge potential but is often considered to be under-used in education research, arguably due to the potential pitfalls that have been identified with this approach and a lack of quantitative skills in researchers (Smith, 2008). Nonetheless, secondary analysis offers a number of strengths: research can be conducted on a large scale that is not often possible when collecting primary data; data can be analysed and interpreted from a number of perspectives; datasets are typically available for free or for a nominal fee; the data is typically high quality; and secondary analysis can inform the collection of primary data later in the research (Gorard, 2002; Smith, 2008). Furthermore, secondary data analysis is unobtrusive and reduces some of the ethical issues associated with using primary data (Smith, 2008).

The quantitative phases of this research focused on the secondary analysis of two datasets: national data from the exam boards and the NPD. Such datasets lend themselves to the analyses in this research, since they provide a complete set of data on national examination entries and outcomes. This allowed the research to explore the research questions and draw conclusions at the national level. Secondary analysis can also set the context for primary data collection (Gorard, 2002), and this combination of

methods is considered a particular strength (Smith, 2008). Thus, in this research the analyses reported in the first phase informed the primary data collection in the second phase, in terms of sampling participants and the interview schedule (see Chapter 5). This adds further benefits to those outlined above of using a mixed methods approach.

While national datasets were used for the majority of the analyses, on some occasions the analysis was restricted to data from a single awarding body, since the national datasets were not sufficiently detailed or nuanced. For example, in GCSE mathematics the national datasets do not specify the tier that students certificate at (see Chapter 1 for a discussion of tiering), meaning that analyses of any trends in tier level entries at the national level are not possible. Instead, data were used from a single awarding body to explore such trends. While it cannot be assumed that the students certificating with this single awarding body are representative of the national cohort, these analyses allowed further insight and additional questions to be explored than previously, hence were considered worthwhile despite these limitations.

Despite the strengths of using secondary data analysis, there are also a number of issues that require consideration. Secondary data is often criticised for containing errors or misclassifications (Smith, 2008), and this might limit its use. However, it is argued that no data can ever be error free and that it is the responsibility of the researcher to acknowledge the limitations of the data and approach its use with caution (Smith, 2008) – as one would when using primary data. Thus, in the present study, appropriate data cleaning was required, decisions were undertaken about how best to use certain variables, and certain individuals were excluded from the dataset (see Chapters 4 and 6). However, this was not necessarily more time consuming than if one were analysing data collected through surveys, for example (Gorard, 2002).

Secondary datasets can also be criticised for excluding certain individuals or types of individuals (Smith, 2008). This is relevant in the present research, since some of the school level data in the first set of quantitative analyses excluded a number of

independent and selective schools (see Chapter 4). Thus, it is noted that a subset of the findings are unlikely to be applicable to this small group of schools. While this might be considered a limitation it is also worth noting that any research can routinely exclude certain individuals (Smith, 2008). Consequently this, and the other limitations discussed above, are not considered to outweigh the benefits of undertaking secondary data analysis in the present research.

The use of interviews

The qualitative phase of this research (phase two) used semi-structured interviews with individuals based in secondary schools in England. Interviews are commonly used in qualitative research (Bryman, 2012), yet the challenges of conducting interviews should not be underestimated (Fontana and Frey, 2005; Kvale, 1996; Roulston, deMarrais and Lewis, 2003). For example, respondents can behave in unexpected ways; the focus of the interview can be lost through poor phrasing of questions; the interviewer can be distracted by participants' responses, the interview setting, or by not listening attentively; and the interviewer can struggle to construct questions that are sufficiently open-ended and relate to the topic (Roulston, deMarrais and Lewis, 2003). Novice researchers may also be challenged when required to follow up responses and clarify meaning with participants, instead moving onto the next question or topic on the schedule (Roulston, deMarrais and Lewis, 2003).

Kvale (1996) identifies 10 criteria for a successful interviewer, that the researchers in the present study sought to adhere to: the interviewer should be knowledgeable of the topic; provide structure for the conversation; steer the interview to focus on what the interviewer wants to find out; ask clear questions; give respondents time to finish what they are saying and think; listen attentively and respond empathetically; be flexible and responsive to what is important to the interviewees; be prepared to challenge respondents (e.g. for inconsistencies); remember previous points and relate information; and clarify interviewees responses without imposing meaning. While the researchers in the current

study sought to adhere to these principles, they are not without challenge. For example, the interviews in this study involved in-depth discussions with participants that focused on specific details of each school's approach to entering assessments. Remembering such details and not confusing them between interviews proved challenging, particularly when several interviews were conducted in quick succession.

Another possible issue stemmed from the use of an additional assistant researcher to conduct two of the interviews, primarily due to practical constraints and the location of the main researcher and the participating schools (see Chapter 5). To conduct a successful interview the assistant researcher was required to develop her understanding of the topic and the interview schedule, despite not having been involved in the research prior to this. This might have led the assistant researcher to adhere more strictly to the interview schedule than the main researcher, since they were less knowledgeable about the topic.

While interviews were the preferred method in this study, alternative qualitative methods such as focus groups was also considered. Focus groups have been defined as '*a research technique that collects data through group interaction on a topic determined by the researcher*' (Morgan, 1996). In this study, focus groups might have allowed data from a greater number of participants to be collected (and indeed more quickly), since multiple individuals from different schools could be accessed at one time (speaking to multiple individuals from the same school would have likely been less useful). However, focus groups would likely have been logistically difficult to co-ordinate between schools and might have limited the depth of discussion, since the questions were likely to touch on sensitive topics that individuals might be reluctant to discuss freely – either with their own colleagues, or their peers from other schools. This does not necessarily mean that the data collected via a focus group would have been less valid (Morgan, 2001) – rather, that different methods are likely to serve different purposes. For example, in some instances the group dynamics in focus groups might facilitate the collection of data for a particular purpose (Morgan, 2001), yet in the current research this was not considered to be the

case. Primarily, this was due to the sensitivity of the issues under discussion. For example, teachers might not have felt comfortable discussing certain issues in front of their head of department or disagreeing with the approach used in their school, since this might have been perceived to reflect negatively on them.

The use of semi-structured interviews

Interviews can vary according to the level of structure that they entail. At one end of the continuum, structured interviews involve a pre-determined set of questions with limited responses, while at the other end, unstructured interviews are open-ended and in-depth – often described as ethnographic interviewing (Fontana and Frey, 2005). This research used semi-structured interviews: similar questions were asked of each respondent on set topics, but there was a degree of flexibility within the interview schedule so that each interview did not necessarily follow the order of the schedule and digressions were allowed to follow up points made by individual interviewees (Bryman, 2012). Semi-structured interviews were considered most appropriate for this study since although the qualitative phase was exploratory, the research sought to explore specific topics and answer pre-determined research questions (Bryman, 2012). Furthermore, they allowed in-depth discussions with individual participants on what was anticipated to be a complex topic.

Interview transcription

When using any form of interview to collect data, there is some debate around the extent to which interviews should be transcribed and how far the researchers should transcribe the interviews themselves. Walford (2001) argues that there can be no set rules surrounding transcription and that it should depend on the research questions being addressed. In the current research, the interviews were transcribed verbatim by an external transcription company, primarily for practical reasons. This approach might be criticised since it is often argued that transcription should be part of the research process, and that enlisting the services of an individual who is not involved in the research could

lead to inappropriate decisions being made during the transcription process (Poland, 2001). To mitigate this, clear instructions were provided to the transcription company concerning the requirements of the transcripts and checks made to ensure that these were followed – for example, the transcribers were advised not to 'tidy up' transcripts, or remove instances where individuals spoke over one another (Poland, 2001).

When considering the transcription of interviews, the researcher also ascribes to the views of Walford (2001), who argues that the act of transcription does not necessarily bring about all of the benefits it purports to achieve. For example, during transcription the researcher can become pre-occupied with the spoken words and phrases, meaning that little understanding of the content is achieved. This suggests that using the written transcripts alongside the audio recordings throughout the data analysis, as in the present study, is likely to offer greater benefits – for example, it allows the researchers to attend to what was being said, and the way it was said (Bryman, 2012).

Ethical considerations and reflexivity statement

Although this research was conducted in accordance with a number of ethical guidelines – including the CERP (AQA) Research Code of Practice (2013), the University of Oxford's ethical guidelines, and the BERA ethical guidelines for Educational Research (2011) – the methods and approaches outlined above brought about a number of ethical challenges and issues that require consideration. These issues are particularly pertinent for the qualitative phase of the research and largely stem from the position of the main researcher as an awarding body employee at the time the qualitative data was collected, and the sponsorship of the research by an awarding body.[41]

[41] The researcher was also employed and sponsored by Ofqual, the awarding body regulator, during the latter stages of this research. However, this was after all of the data collection and analysis had taken place.

The following discussion explores how these affiliations created challenges and ethical issues at various stages of the research, focusing on the qualitative strand. Moreover, it allows me to provide a reflexive account of how my position and values are likely to have influenced the research (e.g. see Mauthner and Doucet, 2003). While recognising how the views and values of the researcher can impact on research and the importance of being reflexive is typically acknowledged (Fox and Allan, 2014; Mauthner, Parry and Backett-Milburn, 1998), the practicalities, complexities, and challenges of doing so are typically less well understood (Finlay, 2002a; Henderson, Holland, McGrellis, Sharpe and Thomson, 2012; Mauthner and Doucet, 2003), and the merits of doing so have been challenged (e.g. see Patai, 1994). However, during this discussion I subscribe to the views of Pillow (2003), believing that while reflexivity does not nullify the issues or necessarily lead to 'better' research, the importance of engaging with, and discussing such issues, remains. To some extent, the ability to be reflexive, particularly when one is conducting research, might have limitations (Mauthner and Doucet, 2003). Thus, while I have made every effort to consider my own influence on each stage of the research, I acknowledge that there may be limits to my thinking and influences that have not been explicitly identified here.

The power relationships in the research

One of the primary issues in the current research stems from the position of the researcher as an awarding body employee during the earlier stages of this research and the implications of this for the qualitative study. While the merits and challenges of a qualitative approach have been discussed above, the use of interviews as a research method brings about a number of ethical challenges. Interviews are sometimes described as a dialogue, yet this notion has been increasingly challenged since it implies that an interview is a joint conversation with equal interest from the parties involved (see Kvale, 2006). Instead, critics argue that interviews entail an asymmetry of power, with the interviewer holding the balance of power over the interviewee (Kvale, 2006;

Vahasantanen and Saarinen, 2012). Kvale (2006) describes a number of ways in which this might happen: the interviewer dictates the content, time, and place of the interview (see also Walford, 2001); controls a one-way dialogue; and holds a monopoly over the interpretation of the data. In the present study, issues of power are particularly pertinent since the researcher was an awarding body employee, and also a student from an academic institution. While it is argued that the latter can instil a level of power in the researcher (Corbin and Morse, 2003), it is the employment of the researcher that is the main focus here.

The employment of the researcher within an awarding body meant that relationships were likely to have existed between the parties involved in the qualitative study. While power relationships between individuals were likely at play during the interviews, relationships also existed at the organisational level between the interviewer and interviewee's respective employers prior to the research being conducted. For example, some respondents were 'customers' of the researcher's employers: they entered their students to the awarding bodies' examinations, and relied upon the awarding body to provide accurate results for their students. Such relationships might have tipped the balance of power further in favour of the researcher, since they could have been perceived by respondents as an authoritative figure. On the surface, this might seem counter-intuitive since ordinarily a 'customer' would be perceived to hold the balance of power over the provider. However, this highlights the complexity of the environment in which the research was conducted, since it is likely that the awarding body (as the provider) also held some controlling powers over the customers, particularly since some of the issues discussed were potentially deemed detrimental to students. While no research could be found exploring the relationship between schools and awarding organisations, there is some evidence to suggest that, within the wider public at least, awarding organisations can be viewed as 'authoritative' (see Chamberlain, 2013).

The complex relationships that existed in this research were likely to have influenced the interactions during the interview. Furthermore, it cannot be assumed that interviewees will always respond truthfully (Walford, 2001). Kvale (2006) discusses how respondents might respond in a manner that they perceive will align with the expectations of the researcher (see also Cheek, 2005) – described as socially desirable response bias. This highlights the importance of considering how to present oneself when conducting interviews, since this can have a large influence over how successful the research is (Fontana and Frey, 2005). While I took care to present myself as a DPhil student collecting data for this purpose, it was also important for respondents to be made fully aware that I was employed by the sponsoring awarding body. It would seem that this led to some respondents viewing me as an awarding body representative. This perception might have inhibited respondents from responding openly and honestly, particularly if they considered that their responses might impact negatively on themselves or their institution. This is a particular issue in the current research given that the behaviours under discussion (i.e. repeated examination entry, examination training etc.) could be construed as unethical or negative, a viewpoint that had frequently appeared in media debate (see Paton, 2013; Sellgren, 2013), and with which the respondents were likely familiar.

Attempts were made to mitigate these issues and encourage participants to respond openly and freely, yet it is acknowledged that such issues are likely to remain inherent in the data. For example, potential participants were informed that participation was voluntary and that they had the right to withdraw at any time, without giving reason and with no negative consequences. Furthermore, they were assured that any data relating to them or their school or college was confidential and would be held securely, and that when reporting any results only high level characteristics of each institution would be reported to ensure that no individuals or institutions were identifiable. There is some evidence that this approach was successful and that participants did respond openly and truthfully. For example, a number of respondents evaluated their schools' approaches

during the interview and discussed what they perceived to be the positive or negative effects of this – or reported an emotional reaction to their responses. It seems unlikely that this would have been the case if individuals had not responded honestly. In other interviews, respondents voiced their lack of support for the approach used within their schools and discussed what they perceived to be the negative consequences of certain approaches, thus highlighting the tension that they faced. Again, it seems unlikely that such discussions would have occurred had participants not responded honestly.

While this might support the veracity of the data, it does raise another concern, in that the issues raised, and the questions asked, might have caused distress to participants. To some extent, semi-structured interviews are likely to have reduced this risk, since participants had a greater level of control over the discussions than if the interviews had been highly structured (Corbin and Morse, 2003). However, it may be that, for some respondents, the discussions brought to the fore the potential tensions that they were facing in attempting to balance their desire to ensure that students achieved a threshold grade, with ensuring that they supported effective teaching and learning (see Chapter 5) – issues that they are perhaps not usually required to confront so directly. It is therefore possible that the interviews left respondents questioning their own values and morals; indeed, one respondent described the moral dilemma that they faced in attempting to balance the needs of the students with the needs of the schools, and others described how they welcomed the return to single examination entry since this would support effective teaching and learning and avoid repeated examination entry (see Chapter 5). This, in itself, does not necessarily mean that the interviews were distressing, and it is acknowledged that interviews can also be a positive experience (Kvale, 1983). Indeed, one respondent commented that the interview was somewhat therapeutic and that they had appreciated the opportunity to discuss such issues openly and freely. However, it is acknowledged that the interviews might have left some respondents questioning their own actions.

Returning to the openness and honesty of respondents, there are two other factors worth considering that are likely to have facilitated this: the timing of the interviews, and where responsibility for the behaviours discussed might lie. The interviews for the qualitative strand of this thesis were conducted following accountability changes announced in September 2013 (after which only a student's first entry attempt counts towards performance tables; DfE, 2013a), meaning that much of the interview required participants to reflect on the entry approaches used for students prior to these changes. This might have facilitated an open discussion, since participants were not describing their current approaches for entering students to assessments. Second, during the interviews it became clear that many respondents placed responsibility for their approach elsewhere, primarily with the Government (see Chapter 5). Again, this might have facilitated openness and honesty, since individuals may have felt that they were reporting an approach for which they themselves were not solely, or entirely, responsible for.

Thus far, it has been assumed that the power dynamics in the interviews were weighted in favour of the interviewer, but the relationship is not necessarily always one-sided. Kvale (2006) discusses how respondents can avoid answering the researcher's questions; respond in the manner they perceive aligns with the expectations of the researcher; or begin to question the interview. As previously discussed, the relationship between the interviewer and interviewee was complex in this research. Some of the interviewees were customers of the researcher's employer and it is possible that this might have instilled a level of power in the respondents as well, since the awarding body relied on respondents' custom to sustain their business. Indeed, one respondent openly challenged the behaviour of the awarding bodies during the interview. While this was not a direct attack on my employer it did lead to a shift in power, leaving me in a somewhat difficult position: as a researcher I was keen to adopt a neutral response to such criticisms, while as an awarding body employee my instinct was to be defensive and provide the appropriate

reassurance. To some extent, therefore, the boundaries, relationships and power dynamics appear to have been blurred and changeable throughout the interviews.

The values and position of the researcher

A second issue requiring consideration relates to the values and position of the researcher and the extent to which this might have influenced the research – both in terms of the overall research strategy and the qualitative phase. There is considerable debate about how far research can, and should, be value-free (see Greenbank, 2003). Some authors argue that all research is likely to be influenced by the values and personal profile of the researcher (Silverman, 2011), including the research methods that are selected (Greenbank, 2003). Indeed, many authors argue that research can never be value-free (Boyd, 2000) and that this should be accepted in educational research (Greenbank, 2003). Researchers are therefore encouraged to adopt a reflexive approach, openly discussing how their values have influenced the research (Greenbank, 2003). In qualitative research, the researcher is often accepted as an integral part of the process (Finlay, 2002a; Henderson, Holland, McGrellis, Sharpe and Thomson, 2012): they can influence the data collection and interpretation (Finlay, 2002a), and it is often argued that their subjectivity should be embraced (Finlay, 2002b). Critics, however, cite a lack of objectivity in qualitative research as leading to biased and subjective research (see Kvale, 1994, for a summary and response, in relation to qualitative research interviews).

Approaching this research primarily as a quantitative researcher, my instinct would be to argue that despite my employment and experiences I have approached, and indeed should approach the research, from a value-free perspective (see also Mauthner and Doucet, 2003). For example, I should seek to suspend my values for the purpose of the research (Berg, 2001), and that this is the most desirable approach. However, while attempts were made to minimise bias, particularly in the qualitative study, it is acknowledged that complete neutrality seems unlikely, if not impossible (Greenbank, 2003). My interest in this research was primarily borne out of my experiences of standard

setting in GCSE mathematics through my employment (see Chapter 1). I had observed first-hand changes in entry patterns over a number of years and had speculated on the drivers behind this. Furthermore, I had experienced and been required to deal with the challenges that changes in entry patterns bring to the process of maintaining examination standards (see Chapter 8). These experiences are likely to have influenced various stages of the research, including the overall research questions and the approaches taken. Perhaps most significantly though, they are likely to have influenced the qualitative study, including the development of the interview schedule, the interpretation of the data, and even my attitude towards participants. Accepting and acknowledging that this does not necessarily have a detrimental effect on the research required a shift in my own thinking, which has primarily developed through a quantitative lens, where controlling possible confounds in the data as tightly as possible is generally favoured.

While the researcher's influence can be embraced in qualitative research, steps were taken to ensure that my own experiences and values did not bias this stage of the research in a negative manner. For example, at the data collection stage careful consideration was given to the development of the interview schedule and the questions that were posed to respondents. The interview schedule was semi-structured and contained several sections, allowing the interviewer to ask high level questions at the start of each section before following-up points as necessary. This aimed to ensure that any preconceptions of the researcher did not overly influence the focus or direction of the interview. Despite this, it is acknowledged that some of the theories I held prior to the research were borne out in the qualitative data. For example, the factors influencing teachers' decisions to move students between modular and linear GCSE mathematics specifications had been theorised by myself and colleagues from analysing data for operational activities (see Stockford 2012a, 2012b) – assertions that were then borne out by some respondents in the interviews. While in such instances suggestions were not provided to respondents based on any preconceptions, I did, if necessary, ask follow-up

questions to confirm or disconfirm whether the theories generated from the data were borne out in practice. Asking questions that might be considered ‘leading’ could be construed negatively, yet in this instance the questions were used to seek clarification and assess the reliability of individual’s responses (Kvale, 1994)

During the interviews, I also sought to adopt a neutral stance to avoid projecting my own opinions and perspective on to respondents. However, while I was well versed in the issues discussed from an awarding body perspective, I had significantly less awareness (prior to conducting the research) of any views held within schools or of school based individuals (what could be described as ‘the other side of the story’). Thus, while I attempted to avoid making value judgements or labelling behaviour in certain ways, I did have an opinion about the behaviours being discussed that had developed through my employment. For example, I had experienced first-hand the challenges that EMEE can create for maintaining standards, so if anything my attitude was likely to be somewhat negative. Thus, while attempts were made to remain neutral throughout, it is unlikely that I portrayed an entirely neutral stance.

My own views and experiences were also likely to influence the data analysis and interpretation phase of the qualitative research. Data analysis is often described in a mechanistic way, suggesting a level of neutrality that is unlikely to be achieved in practice (Mauthner and Doucet, 2003). For example, a number of authors have argued that the analysis of qualitative data is not neutral and is influenced by the assumptions of the researcher (Alvesson and Skoldberg, 2000). The researcher makes choices during the analysis: they decide how to interpret the data, and decide what evidence to present from the interviews to support their findings (Mauthner and Doucet, 2003). When conducting this study, I brought an array of prior experiences and preconceptions to the analysis through my employment as a researcher within an awarding body. These preconceptions could cloud the impartiality of the data interpretation and the generation of themes in the data, resulting in particular themes or issues being prioritised. For example, my own

opinions and expectations might have led me to focus on particular aspects of the data and become blinkered, dismissing or simply not observing counter examples to these. This might result in what Kvale (1994) describes as 'biased subjectivity', with the researcher *'only noticing evidence supporting their own opinions, selectively interpreting and reporting statements justifying their own conclusions, overlooking any counterevidence'.* This, in itself, would be problematic, and steps were taken to minimise this: a second researcher[42] independently coded the transcripts line by line, both researchers sought to focus upon the data rather than the questions, and an inductive approach to coding the data was adopted, where the analysis proceeded *'without trying to fit it into a pre-existing coding frame or the researcher's analytical preconceptions*' (Braun and Clarke, 2006). Furthermore, the researchers sought counter examples in the data that did not fit with, or challenged, their emerging themes.

While the values and position of the researcher created a number of challenges for the research, particularly during the qualitative phase, this can also bring strengths to the process. For example, I believe that my increased knowledge (gained through my employment) of the assessment system, GCSE mathematics assessment, and the processes for maintaining standards, has afforded me with far greater insight into the issues discussed than those who are not embedded within the assessment system. Having 'insider' knowledge therefore allowed me to develop the research programme with a high degree of prior knowledge and understanding. Furthermore, in the qualitative phase, I believe it facilitated more in-depth discussions with respondents and allowed specific issues to be raised, and provided me with a greater understanding of the data.

[42] Note that while the second researcher was also employed by the sponsoring awarding body, they were less familiar with EMEE in GCSE mathematics so were unlikely to hold the same preconceptions and expectations as the main researcher.

A final point to consider relates to my identity and how this has evolved throughout the research. While it has been acknowledged that my values and position are likely to have influenced my approach to the research, the research process itself has also shaped my views and identity (see Thomson and Gunter, 2011) – both in relation to research methodology, and the issues that have been discussed. Reflecting upon my own influence on the research has introduced me to a wide range of issues that require consideration, particularly when conducting qualitative research. This has allowed me to appreciate how the influence of the researcher is not necessarily detrimental to the research. Additionally, discussing the issues under consideration with individuals based in schools has introduced me to the opinions and views of other stakeholders who are involved in the assessment process – knowledge that I could not have gained through operational experience alone. Thus, following the research I consider myself to have a greater and more balanced understanding of the issues and, similar to Walford (2001), my initial perspectives were modified to some extent by the research.

Recruitment and incentives

The discussions above have focused upon how the values and position of the researcher might have influenced the research, focusing on the qualitative strand. Such issues are also likely to have impact upon the recruitment for this phase and who agreed to participate. During the recruitment to the qualitative study it was explicitly stated on several occasions that although I was conducting the research for my doctoral studies as a student at the University of Oxford, I was also employed by the sponsoring awarding body. This could have resulted in individuals feeling compelled to participate, particularly if they felt that non-participation would reflect badly on themselves or their institution.

Steps were taken to mitigate this and there is some evidence that the approaches taken were successful. For example, participants were informed that participation was voluntary and that they could withdraw from the study at any point with no negative consequences, and of the 18 schools contacted initially via the telephone, only three responded

positively.[43] On the whole, it therefore seems unlikely that individuals felt compelled to participate. When recruiting the remaining participants, contact was initially made via email through an existing contact in the mathematics community (i.e. not from the researcher) and individuals were asked to contact the researcher if they were interested in participating. Again, it seems unlikely that individuals would have felt compelled to participate due to the position of the researcher since they had no direct contact with the researcher at this stage. However, this does mean that a certain type of individual might have agreed to take part: for example, those that had strong opinions that they wished to share (see Chapter 5).

A second issue concerning recruitment in the qualitative phase and the position of the researcher stems from three of the respondents also being employed by the sponsoring awarding body as assistant or senior examiners. These individuals may have felt compelled to participate if they considered their future employment to be linked to their participation or the outcomes of the interviews – this might have been in addition to any compulsion to participate due to the reasons outlined above. To some extent, concerns regarding future employment seem fairly unlikely, since the researchers operated in a separate part of the organisation to those directly involved with examiner employment and recruitment. Thus, the researchers have no routine contact with the majority of examining personnel,[44] a set-up that the examiners should have been aware of. Furthermore, decisions regarding the future employment of examiners are to some extent formulaic – for example, they rely on the past performance of the examiners and feedback from supervising examiners – an approach that the individuals involved in the study should again have been aware of. It therefore seems unlikely that respondents considered the

[43] While this might seem a low response rate, it is noted that it was not always possible to access the appropriate individuals within the schools, i.e. access was often via a 'gatekeeper' – often the receptionist – or messages were left on answerphones.

[44] With the exception of senior examining personnel where contact occurs during the awarding process.

researchers able to influence their future employment prospects, at least not directly. Furthermore, respondents were approached due to their position in their respective school, and not due to their examining experience.

A final issue in the qualitative phase relates to the use of incentives. It was anticipated that recruitment would prove challenging for the qualitative study since although the research was conducted at what might be a 'relatively' quiet time in schools (i.e. December – February), the targeted respondents were all full-time members of staff in their respective institutions. Furthermore, some respondents had additional roles and responsibilities in addition to their employment (e.g. they were examination board employees), and recruitment to similar studies in the past has proved challenging. Given this, potential participants were offered a £20 voucher for their time and although there was no intention to coerce individuals to participate, an incentive, by definition, could do this.

To minimise any such effects, the reward offered was small – a £20 voucher for up to 45 minutes of their time – meaning that receiving the incentive (or not) was unlikely to have significant implications for individuals. This approach intended to ensure that respondents felt able to refuse participation, despite an incentive being offered. It seems that this was successful since, as discussed above, recruitment proved challenging and the majority of respondents approached the researcher, rather than the researcher approaching them. Furthermore, upon receiving the voucher at the end of the interviews, many respondents were surprised (and pleased) to be given this. This suggests that the incentive had been forgotten and was not a driver behind their agreement to participate.

Sponsorship of the research

A final point requiring consideration relates to the funding of the research. In addition to being employed by an awarding body, I was also initially sponsored for the research by

the same awarding body (and in later stages by the awarding body regulator).[45] While the relationship between myself and my funder is different to that which might be experienced between an academic and an external funding body (e.g. see Cheek, 2005), there are still implications to consider. The research was primarily funded to further my personal and professional development, hence the research topic was neither pre-determined, nor was it integral to securing sponsorship. However, it was necessary for the topic to be relevant and of interest to my employer, to ensure that they too benefited from my research. While I was therefore afforded a high degree of freedom over the initial topic and the direction of the research, it is likely that the path I have chosen to follow has been influenced by my employment and the experiences I have gained though my employer.

Despite the high degree of freedom I have had throughout the research, a tension can occur when the results of the study do not align with the expectations or views of the funder (Cheek, 2005). While I believe that the relationship I have with both sponsoring bodies is different to that which one would have with an external funder, any papers or reports resulting from my work could, and likely would, be linked to the sponsoring body, and could be construed as the opinion of the examination board or the regulator, rather than my own. This can create a tension, particularly since the qualitative research involved discussing sensitive issues and behaviours with participants and any findings I reported could be viewed as my employer endorsing or criticising such behaviours. This might lead to the sponsoring body suppressing findings, or only allowing the research to be disseminated in a controlled manner (Cheek, 2005). Fortunately, this has been a minor issue for me and I have been encouraged to disseminate my work freely, and while I have

[45] When sponsored by the exam board regulator the data collection and analyses were almost complete, so there was very little scope for the focus of the research to change. Thus, the discussion here focuses on my sponsorship by an awarding body.

been aware of the tensions that may occur, I have sought to ensure that this does not influence the way in which I have interpreted or reported my results.

Summary of Chapter 3

In this chapter the overall research design and the rationale for the approach taken has been discussed. This has included a discussion of the benefits and challenges of mixed methods research, the use of secondary data in the quantitative phases, and semi-structured interviews in the qualitative phase. The chapter has also discussed the ethical issues encountered in the research, and considered the position and influences of the researcher. The following chapter discusses the first quantitative phase of the research.

Chapter 4 – Trends in early entry to GCSE

Overview

This chapter focuses on the first phase of this research and addresses RQ1: what are the key trends in early entry to GCSE mathematics assessments between 2007 and 2011? The key trends are outlined by presenting the findings of an empirical analysis of all awarding body data from 2007 to 2011 (an analysis of multiple entry is presented in Chapter 6).

At the highest level, the analyses consider the number of students who entered early and how this has changed over time for a number of GCSE subjects. Following this, more detailed analyses for the subjects where early entry is most prevalent (English and mathematics) consider the type of students that entered early in terms of their prior attainment, their tier of entry (see Chapter 1), and the characteristics of the school that they attended. The analyses also investigate the performance of students who entered early and how this changed over time, as well as how the performance of students that entered early compares with the performance of older students with the same prior attainment.

Introduction

The analyses in this chapter focus on trends in early entry between 2007 and 2011, building on the literature in Chapter 2. Although early entry has been defined in this thesis as instances where students certificate during Year 10 or mid-way through Year 11, this chapter focus on specific instances: when 15-year-old students certificate at the end of Year 10. This is because including several types of early entry in a single analysis would make comparisons over time difficult, one of the key purposes of these analyses. There are two (related) reasons for this. First, the proportion of students that entered early at different time points (e.g. during Year 10 or mid-way through Year 11) might have

changed over time, meaning that for each group the trends might differ. Second, the characteristics of students that entered early at different time points might differ, meaning that there could be different implications of early entry for each group. Collectively, this could mean that any trends in early entry at the overall cohort level might have been driven by changes within one or other of these groups, meaning that we might not be making like-with-like comparisons between years if we were to include all students in any analyses. For example, the attainment of students entering at the end of Year 10 might have declined, while the attainment of those entering mid-way through Year 11 might have increased, yet these changes could cancel one another out if both sets of students were included in our analyses. Thus, by considering just those students that entered at the end of Year 10 we are able to report a more focused analysis of one type of early entry. Furthermore, it could be argued that students certificating at the end of Year 10 are a more extreme example of early entry than those entering mid-way through Year 11.[46] Thus, although the research questions are written in broad terms, they are applied to the specific type of early entry used in this chapter.

The analyses initially focus on high levels trends showing the percentage of early entries for a number of GCSE subjects (focusing on the larger entry EBacc subjects; see Chapter 1), before considering English and mathematics in more detail since early entry was most prevalent in these subjects (see Chapter 2). Thus, although the overall focus of this thesis is on GCSE mathematics (see Chapter 1), this chapter provide additional context to the changes in entry trends occurring between 2007 and 2011. Moreover, the analyses provide further support for the focus on mathematics since it is here where the changes in early entry exhibited the most interesting trends.

[46] Note that very few students entered GCSE mathematics mid-way through Year 10 according to the DfE (2011), hence these students were not the focus of the analyses.

The analyses in this chapter address RQ1: what are the key trends in early entry to GCSE mathematics assessments between 2007 and 2011? Within this over-arching research question a number of subsidiary research questions were also considered. The research questions all relate to the specific type of early entry used in this chapter (students entering early at the end of Year 10).

RQ1a) what are the trends in the number of early entries between 2007 and 2011?

RQ1b) how do students that enter early perform and has this changed over time?

RQ1c) what type of students enter early (in terms of their prior attainment and centre type) and has this changed over time?

RQ1d) how do early entry students perform compared to older students with the same prior attainment?

Method

Data collection

Data was collected from the five UK awarding bodies for all students in England who sat any full course GCSE subject in a June examination series between 2007 and 2011.[47] The five UK awarding bodies include the three English awarding bodies (AQA, Edexcel, and OCR), the Welsh board (WJEC) and the Northern Irish board (CCEA).[48] Data collection was restricted to the June examination series. While certification for English and mathematics was also available in a November series (and others for mathematics) between 2007 and 2011, the majority of entries occurred in the June series and these

[47] Permission was sought from the inter-board Standards and Technical Advisory Group (STAG) to use this data.

[48] At the time these analyses were conducted GCSE qualifications from each of the five UK awarding body were available in England, hence the inclusion of data from all UK awarding bodies.

analyses focused on students entering at the end of Year 10 (which would be in a summer series).

For each student, information was obtained on the number of GCSE subjects taken, their performance in each subject (i.e. their grade at certification), their date of birth,[49] their school or school type (see details below) and, where possible, their prior attainment at KS2 (see details below). Data was available on every GCSE subject offered by the five UK awarding bodies, a total of 31 subjects.[50] In addition to the student level data, further information on students' schools was obtained via the DfE (see details below) and matched to the student level data.

The analysis was restricted to students based in England only, in keeping with the overall focus of this thesis (see Chapter 1). Nonetheless, there were also practical reasons for this. For example, prior attainment data (i.e. KS2 data, see below) and information on school characteristics (see below) was only available for establishments based in England. Thus, to ensure that the same students were included in each section of the analysis, students from Wales and Northern Ireland were excluded. In addition, the analysis was restricted to students studying in schools only (see below for further information on institution types), who comprised the majority of the 15- and 16-year-old entries for the largest GCSE subjects, English and mathematics, between 2007 and 2011.[51] This was due to their being limited data available on the characteristics of

[49] Due to some errors in the data students who were aged under 14 were grouped together with students aged 14. Likewise, students over the age of 19 were grouped together with students aged 19. These students represent a small proportion of the overall entries, however, and are not the focus of analysis throughout this chapter.

[50] Note that some specifications were grouped together. For example, there are multiple design and technology specifications that were considered as one subject. Hence, the number of subjects given here should not be considered representative of the total number of GCSE qualifications available.

[51] For GCSE English and mathematics in 2011, around 98% of the 15- and 16-year-old entries were from schools.

establishments that were not classed as schools and, again, sought to ensure consistency throughout the analyses.

While the majority of the analyses used all awarding body data (and therefore provided evidence of trends at the national level), some sections were restricted to data from one awarding body (AQA). Where appropriate, this is noted in the text and an explanation provided. With the exception of these sections, however, the remaining analyses should be assumed to represent the situation within schools at the national level.

As the trends in early entry were likely to differ by subject (see Gill, 2010), the initial analyses exploring the age profile of each cohort over time focused on those subjects where early entry was considered most likely to occur. This included the mandatory subjects in performance league tables – English and mathematics – as well as other subjects that have generally been prioritised in measures such as the EBacc (see Chapter 1 and DfE, 2010). This included science subjects (combined science, biology, chemistry and physics), modern foreign languages with the largest entries (French and Spanish), geography and history. Following this, further analyses only focused on those subjects where early entry was particularly prevalent and increasing over time – specifically, English and mathematics. This is due to the low proportion of early entries in the other subjects.

Prior attainment data

To obtain a measure of prior attainment, KS2 results (see Chapter 1) were obtained and matched to students' GCSE outcomes via their name and date of birth. KS2 data is used by the awarding bodies when generating statistical predictions to support the setting of grade boundaries (see Chapter 1), and has been identified as a reliable predictor of

performance at KS4 (Eason, 2010). Hence, it was used as a proxy for student ability throughout this analysis.[52]

Data was available detailing each student's performance in the three KS2 assessments that were available at the time (see Chapter 1): English, mathematics and science. The data is recorded as the level that each student achieved in the KS2 tests, alongside a teacher assessment level. These levels can range from 2 to 5, with a higher score representing higher performance in a subject. In keeping with the procedures adopted by awarding bodies when using prior attainment data for generating predictions (see Eason, 2010), KS2 data was only used for students with complete data for all three tests. Furthermore, the KS2 data relating to the test scores was used, rather than the teacher assessment scores. This inevitably limits the number of students matched to their prior attainment (the match rates are discussed later), meaning that unmatched students were excluded from parts of the analysis. These students were likely to be those who had completed a sub-set of the three KS2 tests or who had attended schools that did not sit KS2 assessments, meaning that the same type of students were likely to be missing from the data each year.

For each student a mean KS2 score was generated by averaging scores across the three tests, giving mean values between 2 and 5 (Table 3 provides an example of the KS2 profile of 16-year-old students within schools in England sitting a GCSE in 2011). Students with scores equal to or lower than 2.67 were grouped together due to the low proportion of students with these scores (for example, see Table 3).

[52] KS2 data was used rather than KS3 data since the latter was not available for all years (see Chapters 1 and 2). Furthermore, KS2 results were being used in the statistical predictions used to guide the setting of grade boundaries at the time that these analyses were conducted.

Table 3. *KS2 scores and 2011 profile of 16-year-old students (n = 601,421)*

Mean KS2 score	Cumulative percentage of students
5.00	20.36
4.67	32.86
4.33	47.86
4.00	75.50
3.67	84.69
3.33	91.50
3.00	97.84
<= 2.67	100.00

School level data

To obtain further information on the characteristics of schools that entered students early, the student level data was matched to the type of institution that each student attended (using data held by the awarding bodies) and further information on each school (using data held by the DfE). At the time that these analyses were conducted educational establishments were classified as one of thirteen institution types by the UK awarding bodies, depending upon their type and their nature of control, using information provided for the National Centre Number Register (NCN).[53] This resulted in thirteen classifications of centre type, although there is some overlap between centres of the same type that have different funding (i.e. they are either community, voluntary aided, controlled or foundation schools). Hence, for the purposes of this analysis, these centres were grouped together, resulting in the following ten classifications of establishments; secondary comprehensive, secondary selective, secondary modern, independent and city academy (all classed as schools), further education establishment, sixth form college and tertiary college (all classed as further education centres), other UK centres (including pupil referral

[53] More recently a fourteenth category has been introduced for 'free schools'. Free schools are funded by the Government but not run by the local council, so have greater control over their affairs.

units) and overseas centres. As this analysis focused on schools only (including academies), the following centre types were excluded; colleges, further education establishments, other UK centres and overseas centres.

In addition to school type, further information on each school was obtained from the DfE via the Edubase dataset and performance league table data. Edubase contains up to date descriptive data on all educational establishments in England and Wales. This includes information on special educational needs, the intake of male and female students in each year group, the religious affiliation of the institution and the percentage of pupils receiving FSM (amongst other data). The Edubase dataset only includes data from the most recent academic year for each school. Therefore, matching the awarding body data to school characteristics was only undertaken for the most recent academic year at the time the analyses were conducted, i.e. for schools with students certificating in GCSE English or mathematics in June 2011. Performance table information is also available via the DfE, showing the percentage of students in each institution achieving five grade A* – C GCSEs in the preceding five academic years.

To match the awarding body data to the Edubase dataset and performance table data, the student level data was initially aggregated to school level. This provided a list of all schools that entered students early or entered students at the end of Year 11 for GCSE English and mathematics in June 2011,[54] as well as an indication of the number of entries from each school. Each of these schools was then matched to their type, name, address and LAESTAB[55] number using data held by the awarding bodies. This resulted in a total of 3,829 unique schools with entries in June 2011 from 15 or 16-year-olds in GCSE English

[54] Schools only entering 16-year-olds students were identified to ensure that schools entering no students early were included in the analysis.

[55] The LAESTAB number is a unique seven digit number allocated to all maintained schools. It contains a local authority number allocated by the Office for National Statistics (first three digits) and an establishment number allocated by the DfE (last four digits).

or mathematics. Of these, it was possible to match 3,610 schools with the Edubase data and performance data, via their LAESTAB number. The unmatched schools tended to be small independent bodies with few entries or schools where the Edubase data related to another academic year (e.g. 2009 – 2010).

While this resulted in a high percentage of the initial population of schools being matched to school characteristics and performance data (94.28%), upon further investigation it was apparent that a proportion of the matched schools were missing certain data in Edubase. Importantly, around a quarter of the schools were missing data corresponding to the number of 15-year-old students that they had registered, information that was needed to calculate the percentage of students in each school that were entered early.[56] Following enquiries with the DfE, it was apparent that this information was missing for data protection reasons: the Edubase database masks certain pieces of information for each school, including the number of students in year groups where schools have fewer than five students, as well as the two year groups for each gender in each school that have the lowest number of students. Inevitably, for some schools, this corresponded to the number of 15-year-old students, meaning that complete data was only available for a sample of 2,770 schools (72.34% of the initial number of schools). Fortunately, aggregated data on student numbers by year group was also available via the January 2011 school census and using this additional dataset allowed the missing schools to be included. This resulted in a total of 3,458 schools with complete data, representing 90.31% of the initial population of schools. Consideration of the missing schools revealed that these were all independent schools that tended to have a low number of students, were high performing, and had no students eligible for FSM, thus were likely to be a distinct group of schools

[56] It was necessary to calculate the percentage of students entering early from each school, rather than using the raw entry figures, due to differences in the overall number of students within each school.

compared to the remaining sample. Thus, the findings reported here are unlikely to be generalisable to this small subset of schools.[57]

Results

The number of early entries

To provide an overview of the trends in early entry and how this has changed over time, initial analyses explored the entries by age group nationally between 2007 and 2011 for each of the larger entry GCSE subjects required for the EBacc (see Chapter 1). Entries were expressed as a percentage of the overall cohort, thus ensuring that changes in the number of students entering each subject each year did not confound the analysis.

The entries for English and mathematics were considered first (see Tables 4 and 5). For both subjects, the percentage of entries from 15-year-old students in the June examination series had generally increased over time, thus concurring with the findings in the literature (see Chapter 2). Furthermore, although the percentage of entries from students aged 14 and under had remained small, they had also increased marginally since 2007, most noticeably for GCSE mathematics. These increases were coupled with a decrease in the percentage of entries from 16-year-old students, suggesting a shift towards entering students early, rather than at the end of Year 11.

Although the increasing trends for early entry appeared largely consistent over time, it is notable that there was a drop in the percentage of 15-year-olds entering GCSE English between 2010 and 2011. This decrease was coupled with a fairly significant drop in the overall entry for English, suggesting that it may have stemmed from schools opting to

[57] Note that the analyses were conducted twice, once including all schools with available data and once excluding all independent schools. As the findings were comparable, the analyses for the whole set of schools is reported.

enter certain students to alternative qualifications, such as international GCSEs (see Chapter 2).[58]

Table 4. *GCSE English percentage of entries by age group (n = total entries)*

Age	2007	2008	2009	2010	2011
<=14	0.04	0.04	0.10	0.24	0.29
15	2.66	3.69	6.55	10.39	8.89
16	94.15	92.99	89.75	85.58	86.91
17	2.52	2.59	2.81	2.86	2.85
18	0.41	0.46	0.52	0.61	0.68
>=19	0.22	0.24	0.27	0.32	0.37
n	603,359	592,899	581,789	588,896	541,466

Table 5. *GCSE mathematics percentage of entries by age group (n = total entries)*

Age	2007	2008	2009	2010	2011
<=14	0.40	0.47	1.25	1.88	2.24
15	8.44	8.95	12.15	13.74	17.53
16	87.31	86.84	82.49	80.05	75.90
17	2.85	2.75	2.89	2.92	2.76
18	0.61	0.62	0.71	0.80	0.89
>=19	0.39	0.37	0.52	0.60	0.68
n	654,171	650,304	663,372	670,401	675,505

The entries by age group for the remaining GCSE subjects included in the EBacc are provided in Appendix E (Tables E1 – E8). As illustrated, for the majority of these subjects the number of early entries as a percentage of the overall cohort remained low and stable between 2007 and 2011, with only GCSE geography and history showing a slight trend towards an increase in early entry.[59] Of the other subjects, there was a higher percentage

[58] Assuming that the size of the national cohort was broadly comparable between 2010 and 2011.
[59] Note that for GCSE science any trends are difficult to identify due to shifts in entries between the combined and separate sciences over time, as well as changes to the structure of the qualifications.

of entries from younger students for the modern foreign languages (French and Spanish), although this had remained stable over time and there appears to be no noticeable shift in entry patterns between 2007 and 2011. These trends concur with other findings relating to the prevalence of early entry in other subjects (see DfE, 2011; Ofsted, 2013). The higher percentage of early entries for languages may be due to native speakers sitting the qualification early (see Ofsted, 2013), or due to schools seeking to complete these subjects early in order to focus elsewhere since languages have not been compulsory at KS4 since 2004 (see Vidal Rodeiro, 2009, for a discussion).

The difference in entry patterns between subjects is likely to be a reflection of the perceived importance of different subjects, particularly in relation to performance league tables and accountability measures (see Chapters 1 and 2), as it is the key subjects of English and mathematics where early entry was most prevalent. Despite this, the prevalence of early entry for English and mathematics was not comparable, with mathematics typically having a higher percentage of early entries (see Tables 4 and 5). Given that both subjects share the common requirement for inclusion in performance measures, the explanation for these differences is not immediately obvious. However, there are a number of possibilities. First, these differences might be due to the structure of the qualifications: while GCSE English adopted a linear structure until June 2012, mathematics was available in a modular format for a number of years prior to this. It is possible that a modular structure lends itself more readily to early entry as students sit the qualification in smaller parts. Thus, entering a younger student for individual modules may be more feasible than entering them for the whole assessment.

An alternative explanation is that the nature of the assessment renders mathematics more susceptible to early entry. For example, it is possible that the extended writing required in English is perceived to be less suitable for younger students (see Chapter 2). Furthermore, English has typically involved coursework (or latterly controlled assessment), whereas this was discontinued in mathematics from 2009. It is possible that

completing coursework with younger students is less feasible or that students are not considered to have developed the necessary skills to undertake this early (see Chapter 2). Third, the entry patterns might be related to the previous success (or otherwise) of schools' entry approaches: it is possible that the higher percentage of early entries for mathematics reflects the greater success of students previously entering this subject early, thus encouraging further early entry. Finally, it might be due to the availability of the assessments (see Chapter 1), since GCSE mathematics was generally available in more examination series each academic year (January, March, June and November).

Given that early entry was relatively low in subjects other than English and mathematics, the remainder of this chapter focuses on GCSE English and mathematics, as it is here where the increase in early entry was most prevalent.

The number of early entries by tier

In addition to considering the overall entries by age group, the above analysis was repeated for English and mathematics focusing upon a student's tier of entry (see Chapter 1). At the time of writing a number of GCSE subjects (including English and mathematics) had tiered entry whereby students could enter higher tier examinations that allowed access to the higher grades (A* – E), or foundation tier examinations that allowed access to the lower grades (C – G).[60] Thus, exploring the entry trends by tier provided further insight into the type of students that entered early and adds to the existing literature (see Chapter 2). However, as the inter-awarding body data only contained a student's final grade and not their tier of entry, this analysis was restricted to data from a single awarding body (AQA).[61] Furthermore, this analysis was only possible for linear specifications, as students were able to combine modules of different tiers when certificating in a modular

[60] Until 2008 GCSE mathematics had three tiers of entry: higher, intermediate and foundation.
[61] In June 2011 AQA had approximately 64.2% of the market share for GCSE English and 25.4% for GCSE mathematics (note that this includes entries from the modular mathematics specification).

structure, thus did not certificate at a particular tier. Consequently, there are limitations to the findings, particularly for GCSE mathematics where the entries for the AQA linear specification were relatively small (e.g. in June 2011 the AQA linear entries comprised less than 6% of the overall national cohort). In addition, the sample is unlikely to be representative of the overall national cohort, as certain types of schools or students were likely to choose to enter the AQA linear specification. Nonetheless, it aims to offer some additional insight into early entry trends.

The entries by age group across tiers between 2007 and 2011 for the largest AQA linear GCSE English specification (English A) and the linear AQA mathematics specification (Mathematics A) are provided in Tables 6 and 7,[62] where the number of students is presented as a percentage of the overall entry. As shown, while the entries from 15-year-old students by tier was relatively comparable in 2007 (2008 for mathematics), over time there has been a gradual increase in early entries on the foundation tier relative to the higher tier. For mathematics, there appears to have been a particular shift between 2008 and 2009, although this may be a reflection of schools adjusting their entry policies following the first sitting of a two tiered mathematics GCSE in 2008.

There are likely to be several factors that have influenced these entry patterns at tier level. First, schools may have increasingly entered their weaker students early to allow them multiple attempts at achieving a particular grade, hence the greater proportion of younger students on the foundation tier (see Chapter 2). This may be a result of the pressure on schools to ensure that their students achieved at least a grade C for performance league tables (see Chapters 1 and 2); indeed, schools might enter a wider range of students early at foundation tier, including the more able, with the aim of securing a grade C. Additionally, schools may have entered their borderline grade C students early at

[62] The analysis for mathematics is restricted to 2008-2011, as the specification was three tiered in 2007.

foundation tier, with the intention of re-entering those who achieved a grade C at higher tier in a later examination series. This might seek to ensure that students achieved a grade C, while also allowing access to the higher grades in a later examination series and again, may result from the pressure experienced by schools from performance measures.

Table 6. *AQA GCSE English A percentage of entries by age group and tier (n = total entries)*

Age	2007		2008		2009		2010		2011	
	F	H	F	H	F	H	F	H	F	H
<=14	0.05	0.03	0.06	0.04	0.04	0.08	0.30	0.24	0.51	0.14
15	2.86	2.51	3.96	3.30	6.99	5.57	11.38	8.62	10.86	7.34
16	92.28	96.17	90.55	95.38	86.65	92.86	81.26	89.57	81.15	90.87
17	3.89	1.07	4.33	1.09	4.98	1.23	5.35	1.26	5.59	1.28
18	0.67	0.12	0.81	0.11	0.98	0.15	1.25	0.17	1.39	0.21
>=19	0.25	0.10	0.29	0.08	0.36	0.11	0.47	0.14	0.51	0.16
n	187,679	240,112	179,968	241,184	168,874	237,806	166,348	236,668	147,967	215,615

Note: F refers to the foundation tier and H refers to the higher tier

Table 7. *AQA GCSE Mathematics A percentage of entries by age group and tier (n = total entries)*

Age	2008		2009		2010		2011	
	F	H	F	H	F	H	F	H
<=14	0.51	0.16	2.23	0.32	3.92	0.28	3.21	1.19
15	4.22	3.65	14.07	4.37	15.12	7.51	20.94	9.52
16	83.46	93.14	69.18	91.17	64.53	87.54	61.46	84.51
17	7.87	2.09	9.47	2.34	9.67	2.37	7.66	2.04
18	2.19	0.23	2.67	0.35	3.57	0.53	3.07	0.35
>=19	1.74	0.73	2.38	1.46	3.19	1.77	3.66	2.40
n	27,001	30,424	28,684	22,264	24,619	18,439	23,352	16,698

Note: F refers to the foundation tier and H refers to the higher tier

To summarise, the number of students that entered early as a percentage of the overall cohort increased between 2007 and 2011, particularly for GCSE English and mathematics. These increases were most prevalent on the foundation tier, suggesting that schools were increasingly entering their weaker students early, or were entering stronger students at foundation tier initially with the intention that they would re-sit at higher tier. To consider the type of students that entered early further, the following sections explore the performance of early entry students over time and the type of students that entered early, in terms of their prior attainment and the type of school that they attended.

The performance of early entry students

If, as suggested above, schools were increasingly entering their weaker students early, it is likely that this will be reflected in the performance outcomes of early entry students. There is some evidence in the literature to suggest that the performance of early entrants declined over time (see Chapter 2). This analysis adds to this and provides an analysis over a longer time period.

To explore the performance of early entrants, the cumulative percentage outcomes for 15-year-old students sitting GCSE English and mathematics across all awarding bodies in the June examination series between 2007 and 2011 were calculated. While this does not account for any differences in the prior attainment profile of the students each year, it aims to provide an indication of whether the type of students entering early had changed during this period.[63]

Tables 8 and 9 summarise the cumulative percentage outcomes for 15-year-old students at subject level for GCSE English and mathematics, respectively, between 2007 and 2011. As illustrated, for both subjects the percentage of 15-year-old students achieving

[63] Note that any comparisons of the outcomes across years assumes that standards are being maintained over time (see Chapters 1 and 8).

the highest grades (A* – B) steadily declined between 2007 and 2011, while the number of entries from this age group continued to rise (see Tables 4 and 5). For mathematics, there was also a decline in the percentage of early entries achieving a grade C or above (Table 9), although for English the outcomes remained relatively stable at this grade (Table 8). The decline in outcomes at the higher grades either suggests that weaker students were increasingly being entered early, hence the decrease in outcomes, or, that there was an underlying factor resulting in this decline in performance. The fall in the percentage of students achieving grades A* – B is likely to be linked, to a certain extent, to the increasing proportion of early entries on the foundation tier, as these students were unable to achieve above a grade C.

While the number of early entries for GCSE geography and history only increased marginally over time (see Appendix E, Tables E1 and E2), a similar decline in performance was observed (see Appendix E, Tables E9 and E10). Here, the percentage of 15-year-old students achieving grades A* – C decreased from 78.97% and 73.43% in 2007, to 65.34% and 62.68% in 2011, respectively. One explanation for this decline in performance across subjects is that there was a shift in the type of students that entered early, in terms of their ability and the schools that they attended. The following sections explore this further for English and mathematics, first by considering the prior attainment of students that entered early.

Table 8. *GCSE English cumulative percentage outcomes for 15-year-olds*

Year	A*	A	B	C	D	E	F	G	n
2007	3.89	15.44	33.69	56.62	75.12	87.23	93.83	97.22	16,032
2008	3.57	14.14	32.97	57.63	76.46	87.47	94.05	97.46	21,853
2009	2.23	9.45	25.54	52.56	75.77	87.76	93.85	97.19	38,136
2010	1.79	9.21	25.52	55.43	78.88	89.86	95.03	97.71	61,186
2011	2.31	9.74	25.88	54.11	77.90	89.62	95.10	97.69	48,150

Table 9. *GCSE mathematics cumulative percentage outcomes for 15-year-olds*

Year	A*	A	B	C	D	E	F	G	n
2007	7.17	23.39	39.86	65.68	78.76	86.27	91.83	95.48	55,211
2008	7.73	23.15	39.80	65.49	79.49	87.04	92.49	96.32	58,197
2009	5.97	19.95	34.83	60.35	76.10	85.31	91.94	96.50	80,575
2010	6.02	18.43	32.48	58.35	74.64	84.92	92.51	97.21	92,124
2011	4.93	15.30	28.57	54.64	71.80	83.56	92.40	97.61	118,425

The prior attainment of early entries

One feature that may explain the declining performance of early entry students is a change in their ability. Throughout this chapter, prior attainment at KS2 (a student's mean performance across the three KS2 assessments in English, mathematics and science) has been used as a proxy for ability. As such, in this section, the analysis was restricted to those 15-year-old students certificating in GCSE English and mathematics that were matched to their prior attainment. As shown in Table 10, this represented a high percentage of the overall 15-year-old entry each year; hence, the final sample was considered broadly representative of the overall entry for this age group. Furthermore, given that the match rates were similar across years and subjects, it is anticipated that the matched students were comparable from one year to the next, in the sense that those missing from the analysis each year were missing for the same reasons. For example, they did not sit all three KS2 assessments (most likely through absence) or they attended schools that did not sit KS2 assessments at all.

Table 10. *Percentage of 15-year-old GCSE students matched to prior KS2 attainment*

	English		Mathematics	
Year	Matched %	n	Matched %	n
2007	80.14	16,032	83.98	55,211
2008	81.80	21,853	84.65	58,197
2009	83.47	38,136	84.84	80,575
2010	84.62	61,186	84.61	92,124
2011	83.40	48,150	84.97	118,425

Figures 5 and 6 show the distribution of 15-year-old students according to their mean KS2 score between 2007 and 2011, for GCSE English and mathematics, respectively.[64] The change in the KS2 profile of students entering early is perhaps most straightforward to interpret for mathematics. Here, there was a gradual decrease over time in the percentage of students entering early with the highest KS2 mean scores (KS2 mean = 4.67 and 5.00). These declines are (inevitably) offset by an increase in the percentage of students with mid and lower range KS2 scores, thus suggesting a shift in the ability profile of early entrants over time. For English, identifying any trends is less straightforward though, with non-uniform fluctuations between years. For example, while the percentage of students with the highest KS2 score (KS2 mean = 5.00) was similar in 2007 and 2008, there was a dip in 2009 before it increased again, then remained stable between 2010 and 2011.

Despite this, these trends are likely to provide some explanation for the declining outcomes for students who entered early, and suggest a shift in schools' entry policies towards entering more weaker students early (see also Chapter 2). Furthermore, they also suggest that there was a greater shift in the entry profile (in terms of the prior attainment of the students) for mathematics compared to English. This would concur with the analyses exploring the outcomes for 15-year-old students over time, where mathematics saw the greatest changes in outcomes, particularly at grade C (see Tables 8 and 9). These differences could be a reflection of the entry approaches adopted by schools for different subjects, or, indeed, the effectiveness of such approaches.

[64] Note that this does not control for any inflation or deflation in KS2 results between years (see Eason, 2010, for a discussion). However, any such changes during the time period that these analyses relate to would be relatively minor and not likely to influence the overall findings.

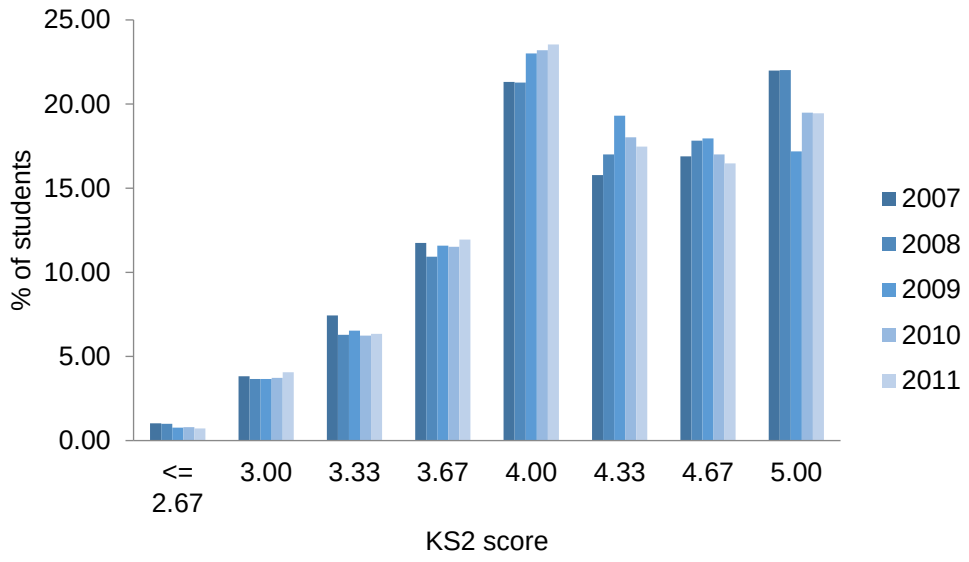


Figure 5. *KS2 profile of 15-year-old GCSE English entries between 2007 and 2011*

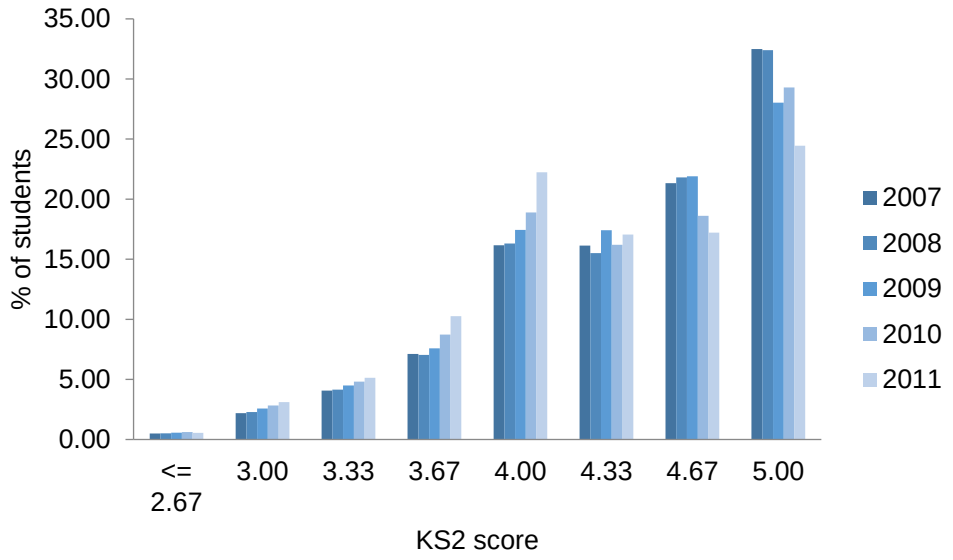


Figure 6. *KS2 profile of 15-year-old GCSE mathematics entries between 2007 and 2011*

Early entries by school type

The decline in the prior attainment of early entry students suggests that there may have been a shift in the type of students that entered early, particularly for mathematics. To

understand this further, consideration was given to the characteristics of schools that entered students early. This analysis replicates and extends previous findings (see Chapter 2), by providing an overview of trends at the school level over time. First, school type was considered. This is based on categories used by the awarding bodies (see earlier discussion in this chapter) and offers a relatively crude method of classifying schools into different types. Therefore, to explore school characteristics further, additional information on each school was obtained from the DfE via the Edubase dataset (see earlier discussion in this chapter). While this dataset included a high proportion of the schools with early entries in June 2011, the findings might not generalise to the small subset of independent schools that were excluded from the analysis due to missing data.

To explore the entries by school type and how this might have changed over time, the number of 15-year-old students as a percentage of the overall entry for each school type was calculated for GCSE English and mathematics between 2007 and 2011 (Tables 11 and 12). This aimed to consider whether, over time, there has been a shift in the percentage of students entering early within certain school types. As shown in Tables 11 and 12, this does seem to have been the case. For both English and mathematics there was a steady increase in the percentage of early entries from comprehensive schools, secondary modern schools and academies. In contrast, the percentage of early entries from independent schools tended to remain relatively stable over time, or increased at a slower rate.

Further analyses shown in Figure 7 offer additional support for these findings. Figure 7 shows the mean percentage of students that individual schools in each category entered early (in contrast to Tables 11 and 12 that consider the whole population of students in each school type collectively). Figure 7 shows that independent schools, on average, entered the lowest percentage of their students early in June 2011. These findings concur with those reported by the DfE (2011; see Chapter 2) relating to the previous academic year.

Table 11. *GCSE English % of 15-year-old entries by school type (n = total entries)*

Year	Comprehensive	Selective	Secondary Modern	Independent	Academy	n
2007	2.70	2.85	2.80	1.96	2.73	16,032
2008	3.53	4.16	5.66	2.94	3.97	21,853
2009	6.82	2.95	9.52	3.43	6.74	38,136
2010	10.80	6.33	15.17	3.70	11.09	61,186
2011	9.03	7.41	12.59	2.49	10.05	48,150

Table 12. *GCSE mathematics % of 15-year-old entries by school type (n = total entries)*

Year	Comprehensive	Selective	Secondary Modern	Independent	Academy	n
2007	8.43	13.39	11.01	5.36	8.44	55,211
2008	9.00	13.40	11.03	6.87	8.67	58,197
2009	11.80	14.71	15.29	8.40	13.01	80,575
2010	13.50	16.22	17.55	7.78	14.65	92,124
2011	17.34	15.47	21.50	9.56	18.93	118,425

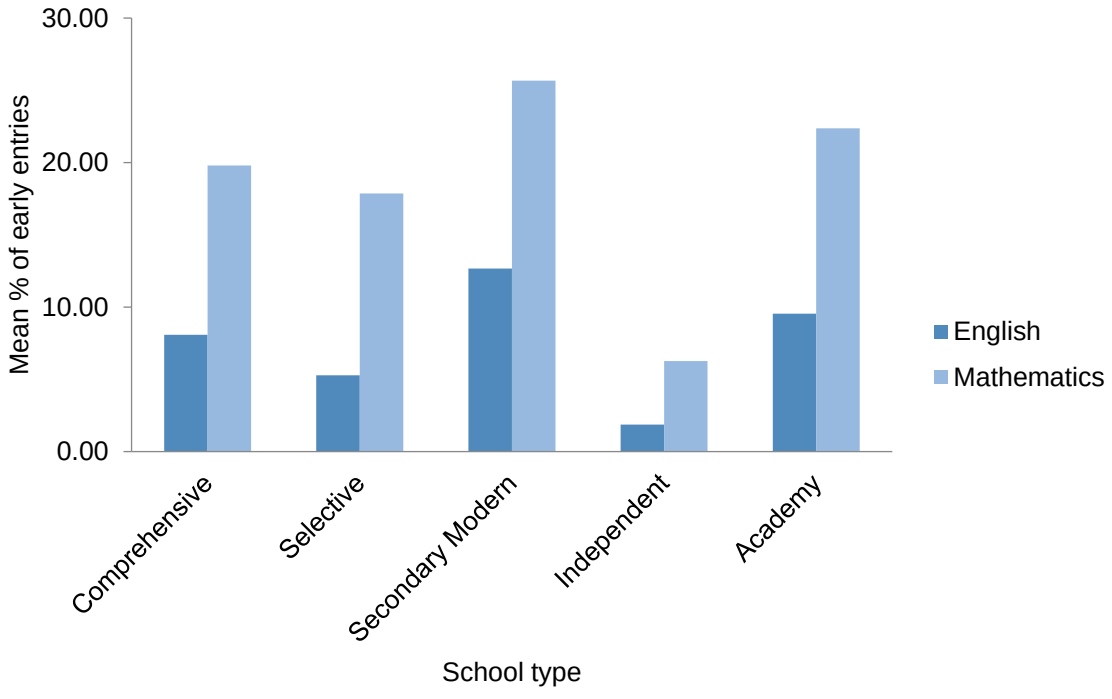


Figure 7. *Mean % of early entries for English and mathematics by school type*

While there appears to be different entry patterns amongst school types, it is anticipated that the characteristics of the schools within each group are also likely to differ. Therefore, further analyses considered the characteristics of schools that entered students early, with the aim of replicating and updating the findings reported by the DfE (2011; see Chapter 2). First, consideration was given to the percentage of 15-year-old students who were entered early, according to the percentage of students eligible for FSM in each school. FSM is often considered as a measure of deprivation in education studies, although the appropriateness of this has been debated (e.g. Kounali, Robinson, Goldstein and Lauder, 2008). Despite this, there is evidence that, on average, students eligible for FSM perform worse than students not eligible for FSM (DfE, 2011).

The findings from earlier analyses reported by the DfE (2011; see Chapter 2) suggested that schools with a higher percentage of students eligible for FSM tended, on average, to enter a greater proportion of their students early for GCSE mathematics in 2010. The current analysis updated these findings by using data from the following academic year (June 2011) and extended the analysis, by additionally considering GCSE English.[65]

Figure 8 indicates that, as previously reported, there is a trend towards schools with a higher percentage of students eligible for FSM, on average, entering a greater percentage of students early for GCSE mathematics. Furthermore, this analysis suggested a similar effect for GCSE English. Despite this, there were noticeable differences between subjects with schools, on average, entering a greater percentage of students early for mathematics than English, regardless of FSM eligibility (see earlier discussion in this chapter for possible explanations for this). Nonetheless, these trends suggest that it is likely that the lower performing schools entered more students early, based on evidence that FSM

[65] Schools were grouped according to the percentage of students eligible for FSM using, for consistency, the categories reported by the DfE.

eligibility tends, on average, to be associated with poorer performance outcomes (see DfE, 2011).

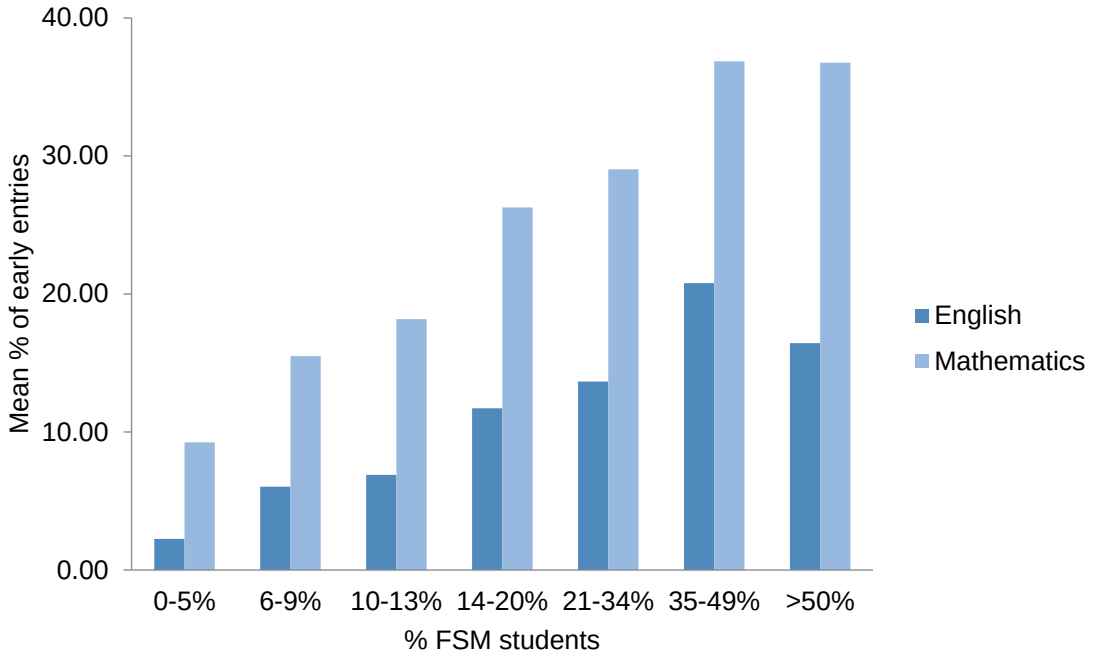


Figure 8. *Mean percentage of early entries for English and mathematics by FSM eligibility*

To explore this further, consideration was given to the performance of schools that entered students early, as measured by the percentage of students within each school achieving five grade A* – C GCSEs including English and mathematics (the key measure used in performance measures at the time; see Chapter 1). Schools were grouped according to their performance outcomes in June 2011[66] and the mean percentage of early entries calculated for each group. This revealed that, with the exception of the lowest performing schools, on average, the higher attaining schools entered a lower percentage of their students early for both GCSE English and mathematics in June 2011 (see Figure 9), thus supporting previous findings (see Chapter 2). An exception to this trend was the

[66] As there was no reason to suggest particular groupings, the following were used: 0-20%, 21-40%, 41-60%, 61-80%, 81-100%.

group containing the lowest attaining schools, who appeared to enter a relatively low percentage of students early.[67] However, this might be due to these schools having fewer students reach GCSE level, regardless of their age. Indeed, of these schools with entry information available for 16-year-old students, the mean percentage of this age group entering GCSE English or mathematics was relatively low for both subjects (34.74% for English and 50.88% for mathematics). However, due to the low number of schools (only 13 of the 16 schools had complete data), this explanation is merely speculative. Interestingly though, if it were the case that lower attaining schools generally had fewer students reaching GCSE level, then the higher percentage of entries from the lower attaining schools could be considered even more surprising.

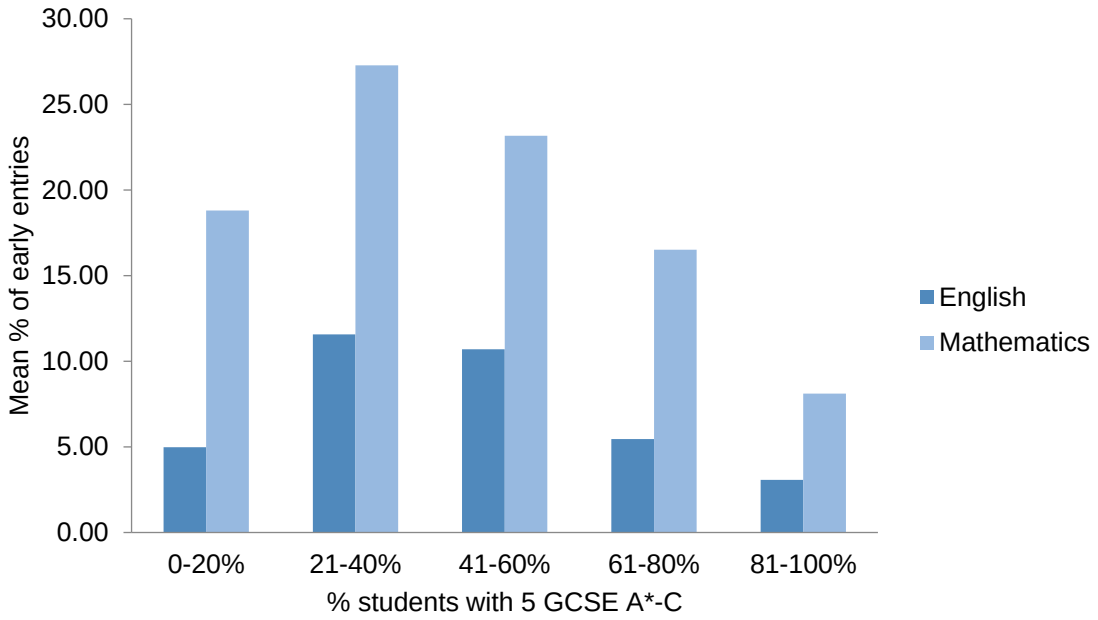


Figure 9. *Mean % of early entries for English and mathematics by school performance*

[67] Note that there were only 16 schools in this group.

The performance of early entries while accounting for prior attainment

Thus far, this chapter has explored trends in early entry with the intention of identifying changes in the number and performance of early entry students over time, as well the type of students that entered early. This final section explored one of the potential effects of early entry for students, specifically, the effect it may have on their attainment. At the time these analyses were conducted there were concerns that early entry could result in under-achievement (see Chapter 2), yet there was little empirical evidence for these claims. This changed in subsequent years, as a number of studies were conducted exploring the possible effects of early entry on attainment (see Chapter 2). Thus, the following analyses add to this literature by exploring the performance of students entering early relative to their older peers, while accounting for their prior attainment. Additionally, it explores whether the effects of early entry differ depending on a student's ability. For example, evidence suggests that higher ability students tend to under-achieve if they enter early (though this is also likely to be linked to opportunities to re-sit; see Chapter 2). The analysis was restricted to AQA only data to allow comparisons between students of different ages who were sitting a particular specification.

To identify any differences in performance, the mean total uniform mark (UMS; see Chapter 1) or raw mark achieved by 15- and 16-year-old students that certificated in the main AQA GCSE English and mathematics specifications in June 2011 were calculated for students according to their mean KS2 score (English A, Mathematics A and Mathematics B). A student's total mark was deemed preferable to their grade since this gives a finer level of detail about a student's performance and facilitates comparisons across groups. For example, a student achieving a grade C may only achieve one additional mark to a student achieving a grade D, meaning that any differences between age groups may appear more or less extreme when using the final grade as the outcome.

For each specification the analysis was conducted at tier level.[68] There were two reasons for this; first, in a linear specification, the total marks are not comparable across tiers, and second, even if they were, the value added relationship between KS2 and GCSE cannot be assumed to be the same across tiers. The analysis excluded students with the top three KS2 scores for the foundation tier and the lowest three KS2 scores for the higher tier,[69] due to there being too few students in these categories for some specifications. Furthermore, students with these scores could be considered atypical, since it would be anticipated that the less able students would enter the foundation tier and that the more able students would enter the higher tier.

Figures 10 – 12 show the mean total mark by tier for 15- and 16-year-old students across the three specifications, according to their KS2 score (see Appendix E, Tables E11 – E16, for the means). They suggest that, for all specifications, there is a tendency for the 16-year-olds to out-perform the 15-year-olds across the ability range, thus providing evidence of a difference in performance between the two age groups. Furthermore, they suggest a trend towards there being a greater difference in performance at the lower end of the ability range, where the 16-year-olds out-performed their younger peers to a greater extent.

To explore the effect of age, prior attainment, and the interaction between age and prior attainment on performance, multiple regression analyses were conducted separately for each subject and tier. With the exception of the higher tier of Mathematics A, this revealed a significant effect of age and prior attainment on performance (see Appendix E, Table E17, for a summary of the output). Additionally, the interaction between age and prior

---

[68] For Mathematics B (modular specification) students were able to combine modules of different tiers when certificating, hence do not certificate to a particular tier. However, their tier of entry on the final module dictated the grades that they could achieve, so this was used in this analysis to define their tier of entry. Within each group there are likely to be students who had combined modules of different tiers though.

[69] For Mathematics A, the higher tier analysis excluded students with the lowest four KS2 scores, due to there being an inadequate number of 15-year-olds in the fourth lowest category.

attainment was statistically significant. Examination of Figures 10 – 12 suggests a greater effect of age on performance for those students at the lower end of the ability range. Therefore, entering weaker students early appears to be the most detrimental in terms of how students perform on this particular assessment. This might seem at odds with the literature that showed that early entry seemed to be detrimental for the higher attainers (see Chapter 2), but possible explanations for this are considered in the discussion.

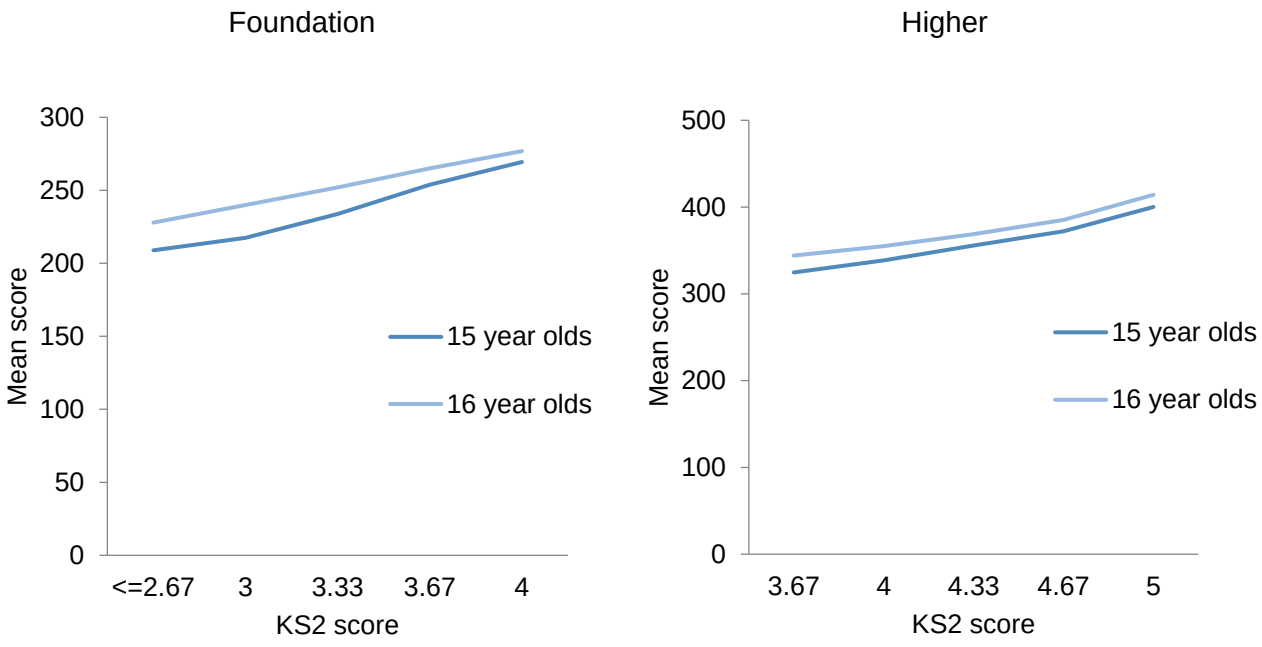


Figure 10. *15- and 16-year-old GCSE English A foundation and higher tier mean mark by KS2 score*

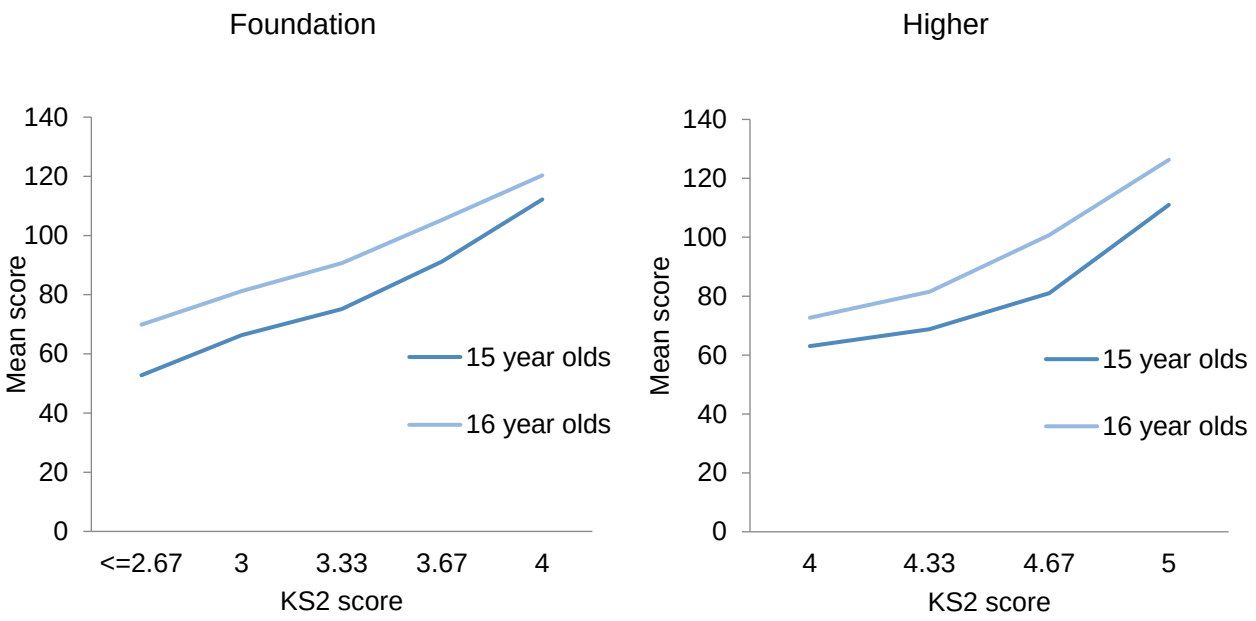


Figure 11. *15- and 16-year-old GCSE Mathematics A foundation and higher tier mean mark by KS2 score*

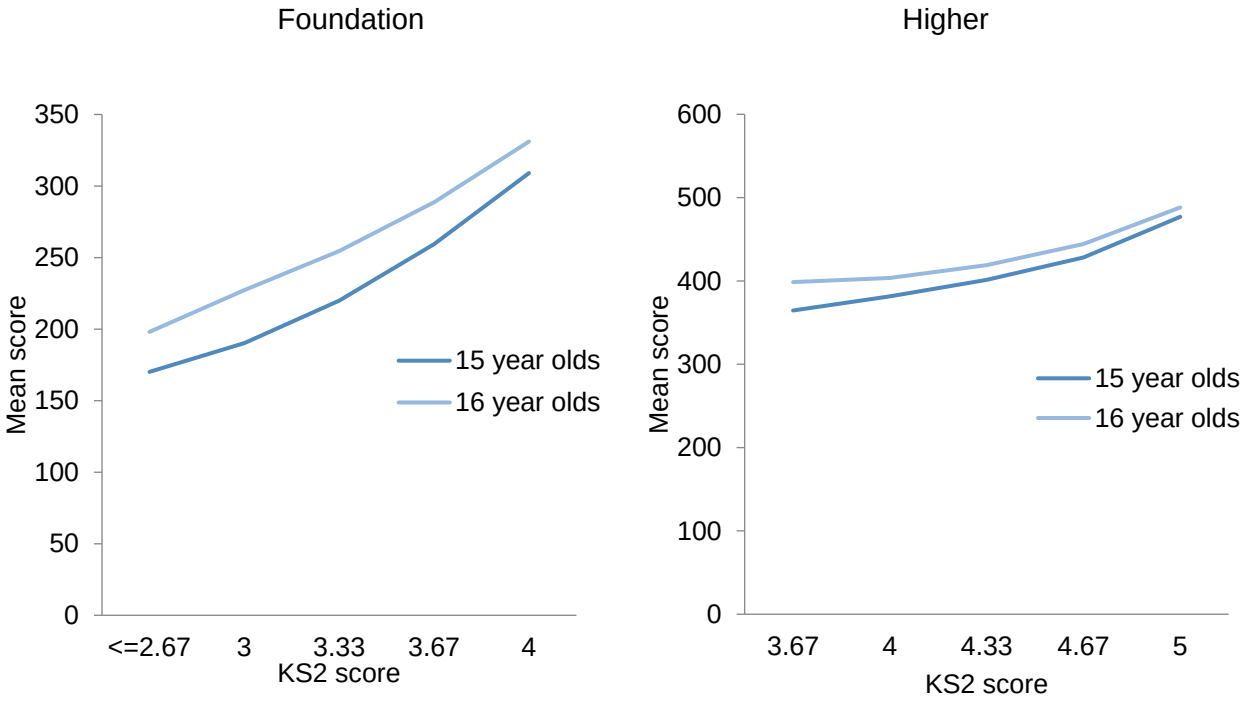


Figure 12. *15- and 16-year-old GCSE Mathematics B foundation and higher tier mean UMS mark by KS2 score*

Discussion and limitations

The findings in this chapter provide additional insight into the presence of early entry in GCSE English and mathematics that builds on the existing literature, yet there are a number of limitations worth noting. First, the analyses considering trends over time at the school and student level might not be generalisable to a small subset of independent schools. This is because these schools were excluded from the analyses due to missing data. Second, these analyses have only focused on a June examination series. It is possible that different schools adopt different entry policies and some, for example, may choose to enter their students early but at a different time point, such as in the November examination series. Nonetheless, the findings support those reported previously (see Chapter 2) and offer an insight into the factors that may influence early entry. For example, performance measures are likely to be particularly pertinent for the lower performing schools (see Chapter 1), suggesting that such measures are likely to have influenced those schools' approaches to entering students to assessments. This may particularly be the case due to the raising of floor targets around the time that these analyses were conducted: in 2010 schools below floor targets were identified as those where fewer than 30% of students achieved five GCSEs at grade A* – C (including English and mathematics), a figure that increased to 35% in 2011 and 40% in 2012 (Acquah, 2013). This is explored in the qualitative phase to gain a greater understanding of the drivers influencing the increase in early entry. It is likely that such policies will differ between schools, meaning that the findings discussed here are also informative for sampling schools in the qualitative study (see Chapter 5). This supports the multiphase mixed methods design (see Chapter 3).

The second set of analyses in this chapter showed differences in performance for 15- and 16-year-old students entering the same GCSE English or mathematics specifications. While the findings suggest that early entry could have a detrimental effect on performance, there are some limitations worth noting. First, it should be considered that

these analyses only compared the performance of students on one particular assessment at one time point. It is possible that students entering early may re-sit and benefit from the experience of entering early, thus may improve their final grade at age 16. Additionally, students entering early may study an enriched curriculum following their early entry, thus may perform better than students who certificated at age 16 in future assessments. These are likely to be explanations for why the findings reported in this chapter do not seem to align with those in the literature (see Chapter 2), where early entry was thought to be detrimental for the higher ability students who do not go on to achieve the top grades (the difference being that we are focusing on one assessment here, and the other analyses may also consider re-sit attempts).

It is also worth considering the educational significance of the findings. For example, on Mathematics B there are 60 uniform marks between each grade boundary, meaning that in the context of awarding grades, the differences between 15- and 16-year-olds observed in these analyses could be considered relatively minor (see Appendix E, Tables E11 – E16, for the mean scores by age). However, the significance of these differences will largely depend upon a student's proximity to the grade thresholds, as here, a difference of a small number of marks could prove significant in terms of the grade awarded.

Finally, it is worth considering the limitations of using KS2 data as a proxy for ability in this context. As the data is supplied to the awarding bodies as an overall level rather than a raw mark, then averaged across the three subjects, there are likely to be differences in the ability of individual students with the same mean score. Indeed, this is why raw marks were prioritised over grades in terms of GCSE outcomes. Unfortunately, at the time these analyses were conducted students' KS2 scores were not available to the awarding bodies, meaning that the levels were used instead. Nonetheless, there is no reason to assume that the distribution of abilities within a particular KS2 category is likely to differ significantly between the two age cohorts. For example, it is unlikely that all the 16-year-olds are at the top of a particular KS2 level, while all the 15-year-olds are at the bottom.

Hence, it seems reasonable to assume that the students with a particular mean KS2 score are of similar ability across the two age cohorts, although this limitation is worth bearing in mind.

Summary of Chapter 4

In summary, the analyses in this chapter have replicated and extended previous findings by exploring the trends in early entry over time for students entering early at the end of Year 10, thus addressing RQ1: what are the key trends in early entry to GCSE mathematics assessments between 2007 and 2011? While this thesis focuses on GCSE mathematics, the analyses reported here have considered a wider range of subjects to provide additional context to the rise in early entry.

The analyses revealed an increasing trend for students to enter GCSE English and mathematics early between 2007 and 2011 (particularly on the foundation tier in mathematics), as well as a slight increase in early entries for other subjects that contributed to the EBacc. The increase in early entry for English and mathematics was coupled with a decline in the achievement of students entering early, as well as a decline in their prior attainment, suggesting that the cohort of students that entered early changed over time. This is supported by evidence showing that it was comprehensive schools, secondary modern schools, and academies that increasingly entered a greater percentage of their students early, with fewer early entries from independent schools. In addition, there was evidence that early entry was more prevalent in schools with a higher percentage of students eligible for FSM and schools with a lower percentage of students achieving five GCSEs at grades A* – C (including English and mathematics). These findings support and extend the existing literature (see Chapter 2), contributing a number of new findings: they provide a more focused analysis of key trends in early entry than has previously been available, focusing on specific instances of early entry when 15-year-olds certificate at the end of Year 10. Additionally, they provide evidence of trends over a

longer time period than has previously been available, and provide evidence that early entry was more prevalent on the foundation tier of assessment, particularly in mathematics.

These findings, particularly those at the school level, suggest that one possible driver of early entry was the pressure on schools due to performance and accountability measures (see Chapters 1 and 2). Such pressures might result in decisions being made by individual schools (in terms of their entry policies) to maximise the chances of students achieving a particular grade, most likely a grade C. Despite this, there was evidence from these analyses that 15-year-old students certificating early performed worse than 16-year-old students certificating at the end of Year 11. This effect tended to be greatest for the lower ability students, which seems at odds with the increase in early entry for weaker students. While this might suggest an effect of early entry on attainment, it is important to note that these comparisons only focused on the outcomes in one specification at a particular time point. Thus, these analyses differ to those discussed in Chapter 2 where students might have entered early but subsequently re-sat and improved their final grade.

While early entry was prevalent in both English and mathematics (the key subjects required for performance league tables), there was a greater percentage of early entries in mathematics than English. This suggests that schools might have adopted different approaches for different subjects, possibly due to the nature of the subject content or the qualification, or the success of previous approaches. Given these differences in entries, the remainder of this thesis focuses on mathematics, since mathematics is likely to be the more extreme example of early (and multiple) entry, and therefore provide the most interesting insight into this approach (see Chapter 1). The following chapter focuses on the qualitative phase of this research that considered the influences on schools' entry approaches to GCSE mathematics, focusing on EMEE.

Chapter 5 – Schools' entry approaches to GCSE mathematics assessments

Overview

This chapter discusses the findings of the qualitative phase of this research that explores schools' approaches to entering students to GCSE mathematics assessments, focusing on EMEE. The use of qualitative methods builds on and complements the findings in Chapter 4, by exploring in greater depth the approaches used by schools, the drivers behind these, and the implications of schools' approaches. This contributes to the mixed methods approach outlined in Chapter 3. In brief, the chapter discusses the sampling, data collection and analysis, and the key themes that emerged from the data. A peer-reviewed paper resulting from this study has been published in the Oxford Review of Education (see Taylor, 2016, and Appendix F).

Introduction

The need and rationale for a qualitative approach to contribute to the study of EMEE has been discussed in previous chapters (see Chapters 2 and 3), but is recapped briefly here. While a number of previous studies have considered trends in EMEE over time, these studies have primarily relied on analyses of large-scale datasets (see Chapters 2 and 4). The benefits of this are obvious, since they provide an overview of entry trends at the highest level, yet there are also limitations since they cannot contribute to our understanding of why such trends occurred. The following study addresses such limitations by using qualitative methods to explore schools' approaches to entering students to GCSE mathematics assessments, focusing on EMEE. This strand of the research therefore addressed RQ2: what influences schools' entry approaches to GCSE mathematics assessments, in particular, to entering students early and on multiple occasions?

The analyses reported in Chapter 4 focused on early entry since at the time those analyses were conducted much of the debate related to early entry. However, as outlined in Chapter 1, the focus shifted in subsequent years to include both early and multiple entry. As such, the interviews in this chapter focused on both approaches to entering students to assessments.

The qualitative phase of this research was planned prior to the announcement in September 2013 that only a student's first entry attempt would count in school performance measures (see Chapter 1). However, the actual interviews were conducted after the changes had been announced. While this meant that the interview schedule had to be adapted to account for the changes, in many respects the timing was quite beneficial, since it allowed respondents to reflect on their use of EMEE prior to the accountability changes, and to consider their approaches in the future.

Methodology

Participants

Data were collected via 14 one-to-one interviews with individuals based in schools across England, with the exception of one interview where there were two participants.[70] This resulted in a total of 15 participants including four heads of department for GCSE mathematics, nine teachers of GCSE mathematics and two examinations officers.[71] In addition to their roles within the school, three participants were also employed by one of the UK awarding organisations as either an assistant or senior examiner. While this might have resulted in some participants having more extensive knowledge of the assessment

[70] Note that this was counted as one interview, since the participants represented the same school and responded to the questions together.

[71] Examinations officers are responsible for the operation of school examinations (both internal and external). The interview with two participants involved a GCSE mathematics teacher and the examinations officer.

system than others, this was not considered problematic for the purpose of this research; indeed, it may have facilitated more in-depth discussions with some participants or provided an additional perspective. The final sample consisted of 10 male and five female participants, representing 14 schools.

Sampling

The sampling strategy was devised with the aim of ensuring that a range of individuals from schools with different characteristics and entry approaches were included in the research, while acknowledging that, due to the exploratory nature of the research (see Chapter 3), a representative sample was not deemed possible or necessary. When generating a list of schools to approach, consideration was given to the following characteristics:

- the percentage of 15-year-old students within each school who entered GCSE mathematics at the end of Year 10 in June 2012;[72]
- the location of the school;
- the performance outcomes for the school in the 2011 – 2012 academic year (as measured by the percentage of students achieving five GCSEs at grades A* – C, including English and mathematics); and
- the percentage of students eligible for FSM in the 2011 – 2012 academic year.

The latter two variables were considered important since the findings in Chapter 4 showed trends in entry patterns based on these criteria; for example, schools with a higher percentage of pupils eligible for FSM and lower performance outcomes tended, on average, to enter a higher proportion of their students early for GCSE mathematics. Thus,

[72] Note that while this aimed to identify schools with different entry patterns, this only included students who were entered in the June examination series. It is likely that students were also entered prior to the end of Year 10 or throughout Year 11, in the November, January or March examination series.

it was anticipated that entry approaches would vary across these dimensions. In addition, recruiting schools across several geographical locations was considered important since it was anticipated that exam entries could, to a certain extent, be dictated at the local authority level.

Based on the above criteria, two lists of schools were generated – those that entered a low percentage of their students early for GCSE mathematics (less than 25%) and those that entered a high percentage of their students early for GCSE mathematics (above 90%), using data from June 2012.[73] Initial recruitment involved contacting schools on these lists in the local area to the two researchers that conducted the interviews (the North West and Surrey). Contact was initially made with the school reception via the telephone, with the aim of speaking to either the head of GCSE mathematics or the examinations officer. Since recruitment via this method proved challenging (only three of 18 schools responded positively and accessing the appropriate individuals within the schools was difficult), additional methods of recruitment were also necessary. First, opportunistic sampling was used via an existing contact in the mathematics community who identified participants through a network of mathematics teachers based in the Midlands. Second, a snowball sampling method was employed, where existing participants recommended colleagues at other schools to approach.

The final sample comprised 14 schools spread across Greater Manchester (three schools), Nottinghamshire (three schools), Leicestershire (two schools), the West Midlands (two schools), Surrey (one school), South Yorkshire (one school), Hertfordshire (one school) and Greater London (one school). As desired, the sample represented a range of schools in terms of their performance outcomes, the percentage of pupils eligible for FSM and the percentage of students who entered early in June 2012 (see

[73] This was the most recent data available at the time that the sample was generated.

Table 13). Thus, while the sampling method deviated from initial intentions (in that opportunistic and snowball sampling were used), the final sample was considered to include a range of schools. While information was not collected on which awarding body each school used for GCSE mathematics (nor was the intention to recruit schools based on the awarding body they used), it was clear from the discussions that, within the final sample, entries had been made with each of the three main awarding bodies in England (AQA, Edexcel and OCR).

Table 13. *Final sample of schools*

School type	% 15yos entering GCSE mathematics June 2012	% five GCSEs A* – C (inc English and mathematics)	% eligible for FSM**
Comprehensive A	98.67	50	30%≤ and <40%
Academy A	23.40	70	10%≤ and <20%
Academy B	*	48	10%≤ and <20%
Academy C	100.00	47	20%≤ and <30%
Academy D	*	56	10%≤ and <20%
Academy E	11.69	56	10%≤ and <20%
Academy F	*	62	10%≤ and <20%
Comprehensive B	5.11	43	20%≤ and <30%
Academy G	30.54	57	10%≤ and <20%
Academy H	0.74	58	10%≤ and <20%
Selective A	0	96	<10%
Comprehensive C	26.67	68	10%≤ and <20%
Academy J	0.47	62	<10%
Academy K	2.82	73	<10%

* Entry figures were not available for these schools.

** Data has been grouped into categories for this variable to ensure that no schools are identifiable

Data collection

All interviews were conducted face-to-face by one of two researchers: the author (12 interviews) and an assistant researcher (two interviews). Two initial interviews were used as pilot interviews, following which minor amendments were made to the interview schedule. However, since this mainly comprised re-ordering the sections in the schedule,

rather than making any significant changes to the questions, these data were included in the analysis.

The interviews were conducted in late 2013 and early 2014. This was after changes to accountability measures had been announced in September 2013 that effectively ceased the use of EMEE (see Chapter 1). Thus, participants were asked to reflect back on their entry approaches before the changes (see Chapter 3 for a discussion of the possible implications of this), and to consider their approach going forwards.

Interviews were conducted in each participants' school and lasted between twenty-five and forty-five minutes. Prior to each interview there was a brief discussion to clarify the meaning of key terms that would be used throughout the interview. As such, early entry was used to refer to instances where a student certificated during Year 10 or mid-way through Year 11, while multiple entry was used to refer to situations where students entered more than once before the end of Year 11: either by using repeated entry (i.e. an early entry followed by a re-sit in a separate examination series) or by entering multiple specifications in the same examination series. Following this, five broad topics were discussed, based on the interview schedule shown in Appendix G:

- entry approaches to GCSE mathematics prior to the accountability changes announced in September 2013, focusing on EMEE;
- entry approaches post-accountability changes;
- the process of deciding entry approaches within the school;
- the drivers behind the chosen approaches and potential consequences; and
- the effect of entry approaches on teaching and how students were prepared for assessments.

Open-ended questions were used at the start of each section to initiate discussions, following which prompts and follow-up questions were used as necessary. Digressions

from the interview schedule were allowed to explore any interesting points or comments raised by the interviewee. All participants granted permission for the interviews to be recorded and were given a £20 voucher for their time (see Chapter 3).

Analysis

All interviews were transcribed verbatim by an external transcription company (see Chapter 3), allowing the analysis to utilise both the written transcripts and the original audio recordings. While in the first instance the analysis relied on the transcripts, the audio recordings were useful in some instances to clarify meaning or to ascertain the tone of a particular comment. Transcripts were available shortly after each interview, allowing the main researcher to interweave the data collection and analysis stages of the research, which can offer several benefits (Miles and Huberman, 1994). For example, knowledge of the existing data can influence and enhance subsequent data collection, by allowing topics to be explored or clarified in greater detail (Miles and Huberman, 1994).

Data were analysed using dominant theme analysis (Maxwell, 2005), a technique that offers a flexible approach to the analysis of qualitative data (Braun and Clarke, 2006). While a number of approaches are possible within this framework, an inductive approach was used, where the coding of the data proceeded "*without trying to fit it into a pre-existing coding frame or the researcher's analytical preconceptions*" (Braun and Clarke, 2006; Leech and Onwuegbuzie, 2007). Despite this, it should be noted that, as stakeholders in the assessment system, the researchers were not without their own prior knowledge or preconceptions, which may have influenced the interpretation and the reporting of the data (see Chapter 3). However, steps were taken to minimise any potential bias during the analysis and to promote "*rigor and trustworthiness of the findings*" (Leech and Onwuegbuzie, 2007). This entailed the data being analysed by two researchers (Leech and Onwuegbuzie, 2007): the author who had conducted the majority

of the interviews (12 interviews) and another researcher who joined the research at the analysis stage.[74]

The analysis comprised three main stages, allowing the researchers to reflect back on the data as the analysis proceeded. During the initial stage of the analysis both researchers independently read the transcripts in-depth (i.e. line by line) while making notes of possible codes in the data and emerging ideas for organising the data. Following this, the researchers met to discuss the codes and agree a series of themes to describe the data. These discussions allowed the codes and resultant themes to be challenged and refined, before the main researcher read the transcripts for a second time, where instances of support and challenge to each theme were identified. A final discussion of the themes was undertaken between the two researchers. Any disagreements were discussed in-depth and agreement sought. Data were analysed initially using paper copies of the transcripts, then using NVivo 10 software (QSR International, 2012). Further details of the coding, including the coding log, are provided in Appendix H.

Results

The results are reported under the key themes identified during the analysis and are presented within the framework shown in Figure 13. This framework was devised to organise the themes that had emerged from the data, rather than the framework being predetermined or used to drive the analysis. It is noted that, due to the complexity of the data, there are several links between the themes (and the initial codes). Indeed the first, and dominant theme, relating to the way that accountability and the pressure to achieve results drives behaviour in schools, is strongly related to the remaining themes, which describe schools' approaches and the possible implications of these. The final two

[74] The researcher who had conducted two of the interviews was not available when the analysis was conducted, hence another researcher assisted with the analysis.

themes, relating to disquiet amongst teachers and the issue of responsibility, are presented outside the framework presented in Figure 13, since they interact with, and permeate through, the preceding themes.

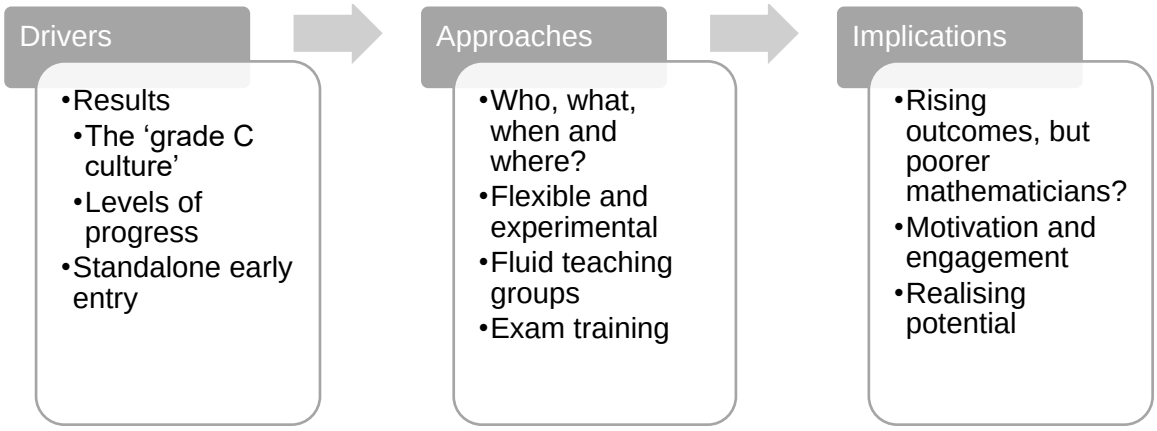


Figure 13. *Framework for key themes*

The results reported here refer to the schools included in this research and, as such, are not necessarily representative of the wider population of schools. Furthermore, no claims concerning the generalisability of the findings are made. However, given the exploratory nature of the research, this is not considered problematic, or unintended.

The results are organised such that a description of each theme is provided, with supporting quotes as appropriate. To ensure that the voices of all participants are heard, quotes are sourced from all of the interviews and the corresponding participant and school is indicated for each.

Drivers

Results – the 'grade C culture'

Throughout the interviews, the key theme emerging from the data related to how the pressure of results and performance measures, particularly around the grade C threshold, influenced schools' behaviour. These pressures influenced schools' behaviour in several ways, both in terms of how students were entered to assessments and the way that

students were taught and prepared for assessments – issues that are discussed in later themes. Essentially, accountability measures led to schools becoming pre-occupied with attempts to maximise their examination results, potentially at the expense of supporting effective teaching and learning. In particular, the pressures around grade C were evident, leading many schools to focus on this threshold. This resonates with much previous research (e.g. Ball, Maguire, Braun, Perryman and Hoskins, 2012; Gillborn and Youdell, 2000; Perryman, Ball, Maguire and Braun, 2011) and in many respects is not surprising, given the importance of performance measures and the potential consequences should schools, and students, not achieve the desired results. At the school level, being deemed as under-performing can result in increased inspections and ultimately closure, while at the student level, not achieving a particular grade can prevent access to further education and employment (see Chapter 1).

> *… you have to think of the context of what you're living in and that was if we didn't hit that floor target, we opened a whole load of hell for ourselves. We opened doors for more Ofsted, more inspections, more LEA inspections and your day to day life was going to be hard …*
>
> *(Head of department, Academy J)*
>
> *… it was like get your C grades or risk shutting.*
>
> *(Teacher, Comprehensive A)*
>
> *… I think the performance tables are everything to a school.*
>
> *(Head of department, Academy C)*

Although results were a key concern for respondents, the degree to which schools felt under pressure differed, with the lower-performing schools tending to report greater pressure to achieve grades than their higher-achieving counterparts (see also Ball, Maguire, Braun, Perryman and Hoskins, 2012; Hutchings, 2015). Thus, the interviewee from a high-achieving selective school reported few concerns about achieving targets, likely due to the high calibre of their student intake. This should not imply that they did not strive to achieve the best results possible; rather, that the pressures from accountability

measures were diminished. This contrast is illustrated in the following quotes, from lower and higher achieving schools, respectively.

> *For me, I believe that what's driven the entry strategy in this school is the pressure from the Government and the LEA on the results of our school. If we didn't step the game up, we would have been put into special measures and they want improvements like that.*
>
> *(Teacher, Comprehensive A) [Low achieving school]*
>
> *There was no rationale. It was merely desperation to get kids at A to C grades …*
>
> *(Head of department, Academy C) [Low achieving school]*
>
> *… we actually got the results up to 70 odd % on a system where the kids just did maths for two years and took the exam at the end of Year 11 so we know we can deliver kids to that position [achieving a grade C].*
>
> *(Teacher, Academy A) [High achieving school]*
>
> *… I mean, to be honest, we could put all the kids in at this school early probably at the end of Year 10 and they'd all get reasonable results.*
>
> *(Teacher, Selective A) [High achieving school]*

For the majority of schools the focus on results manifested itself in a focus on the grade C threshold, one of the key measures in school performance tables (see Chapter 1). This often led to what has been coined an 'exam factory' approach (Mansell, 2007; see also Hutchings, 2015), with schools aiming to get as many students over the grade C threshold as possible, regardless, to a certain extent, of how this was achieved and the implications for teaching and learning (see also Ball, Maguire, Braun, Perryman and Hoskins, 2012; Gillborn and Youdell, 2000; Perryman, Ball, Maguire and Braun, 2011). Students were often discussed as numbers or percentages, an approach facilitated by there being multiple examination series available throughout the academic year that allowed schools to monitor how many students had crossed the grade C threshold as the year progressed. This often reduced the pressure experienced by schools at the end of the academic year,

and resonates with the 'A – C economy' described in earlier work by Gillborn and Youdell (2000).

> *If they get a C, excellent, that's one more off the list.*
>
> *(Teacher, Comprehensive A)*
>
> *… that helped this school as a whole to see exactly what their statistics were in terms of 'ok, now we've got these exam results we know we have got 63% in the bank, our target is 70%, therefore we need another 7%, this equates to 20 more pupils or something so who are these 20 going to be?*
>
> *(Teacher, Academy G)*

In addition, resources were typically focused around C/D borderline students, often to the detriment of other students who had either achieved a grade C already, or were not likely to, an issue discussed later.

> *It might be slightly unfair on students who either we can't see getting anywhere near a C or ones who are comfortably going to get a C because they tend to, you know, they're not a concern …*
>
> *(Head of department, Academy C)*
>
> *… by sort of midway through Year 10, if they're not going to get a C grade, forget about them and no one cared about pulling those kids out of the Es and Fs and turning them into Cs. So if they came in with a target of a C and they were below that, we'd work with them. If their target was, for example, a D, an E or below, then it was a case of 'right, you babysit that group and then we'll look after the other ones'. And similar at the top to be honest with the high achieving kids, 'oh they're definitely going to get a C, oh let's not put too much pressure', it was really focussing on those middle kids and middle achievers, a race to the middle all the time.*
>
> *(Teacher, Comprehensive C)*

The pressure experienced by schools to achieve results, particularly at grade C, was illustrated by the language that respondents used throughout the interviews. Many respondents reported the 'need' for students to achieve a particular grade, either to

ensure that the school achieved a certain percentage of students passing the grade C threshold, or to ensure that students demonstrated a certain level of progress.

> *… that senior teacher would be saying 'right we need to get …', I've mentioned this before, 'if the target's', I don't know, '75%, we've got to get 75%*
>
> *(Teacher, Academy G)*
>
> *… we're measured on how many Cs we got, it didn't matter if they weren't Bs or As and A*s, so you needed to get Cs because if you didn't get enough Cs, you were classed as being a rubbish school …*
>
> *(Head of department, Academy J)*

While for many schools the drive to achieve as many grade Cs as possible was clearly linked to school performance measures and their associated pressures, this was not always the case. Indeed, some schools reported using a particular approach with the aim of ensuring that their students achieved a grade C to maximise their future life chances. Thus, accountability at the school level was not the sole driver. Frequently, this desire to achieve results for the students ran alongside the desire to achieve results for the school.

> *I mean the real drive was to get the best for the kid and the weaker kids, the borderline kids, we just said 'well if you've got the opportunity, give them as many chances as you can.*
>
> *(Teacher, Academy A)*

It is worth considering though that until more recently, what has been in the best interests of the school and borderline students (in terms of achieving a threshold grade, at least) have arguably been closely aligned. Thus, using approaches to maximise achievement at grade C has been beneficial for both the school and the students: a win-win or low risk situation.

> *… the way I've viewed it in previous years is that in one respect it's great for the kids because they get lots of practice but for the school it's almost like they've banked C grades … it's a win/win situation …*
>
> *(Teacher, Comprehensive A)*

However, following changes to accountability measures from September 2013 (see Chapter 1) this was not necessarily the case, since only a student's first result counts in school performance measures. Consequently, maximising the chances of achieving a grade C (through multiple entries) could be beneficial for the student, yet not the school, requiring schools to balance the needs of the school and the students – a tension identified elsewhere (Ball, Maguire, Braun, Perryman and Hoskins, 2012). Striking this balance is likely to be challenging and, amongst the majority of schools included in this research, the needs of the school appeared to take precedence. Hence, nearly all of the schools who had intended to enter students early in November 2013 had withdrawn their entries following the accountability changes, due to the risk that early entry was deemed to pose to results. Furthermore, respondents suggested that, henceforth, they would only enter students once at the end of Year 11, rather than using EMEE.

> *… we don't want them taking it if they're going to get a D, because we need Cs …*
>
> *(Teacher, Academy H)*
>
> *I mean the whole of our Year 11 was entered in November and every single one was withdrawn.*
>
> *(Teacher, Academy A)*
>
> *Yes and for some of those pupils, what's a real shame is it's not in their best interests to wait until the summer, it's in the school's best interest …*
>
> *(Teacher, Selective A)*

Only one respondent reported entering students in the November 2013 series as planned, due to the perceived negative effects that withdrawing their students would have. This

was one of the higher achieving schools and it seems reasonable to speculate that this school may have experienced less pressure to achieve results than their lower achieving counterparts, potentially making their decision more straightforward.

> *One of the reasons we chose not to withdraw is because of the impact it would have had on parents and students because they're all ready for it … if we were to suddenly turn around and say 'sorry we've changed our minds because the Government have changed that', it would have been devastating for the students and the parents would have complained because they saw it as a positive strategy, they felt it was the right thing for the kids …*
>
> *(Examinations officer, Academy K)*

The desire to prioritise the schools' needs above the students is perhaps not surprising, given the potential consequences of schools not achieving the required results, such as increased monitoring, inspections, and, ultimately, closure (see Chapter 1). Furthermore, while there are implications of not achieving a threshold grade for individual students, the consequences for the school are likely to be wider and far reaching; indeed, if the school were to close, there would be significant implications for the students as well.

> *I mean that's not because I don't think this school or any other school doesn't care, I think they do. However, life is so much harder, you indirectly affect your students if you don't have a good … if your school isn't shown to be doing well.*
>
> *(Head of department, Academy C)*

Results – levels of progress

While the discussions throughout the interviews overwhelmingly focused on the pressures around the grade C threshold, it is also worth noting here that another measure, levels of progress, was also discussed by most respondents. This measure, used by Ofsted (see Chapter 1) at the time the interviews were conducted to judge the performance of schools (Ofsted, 2014), considers the number of levels of progress made by students between KS2 and KS4 and was seemingly becoming more important to schools. Thus, schools

were increasingly attempting to strike a balance between ensuring that students achieved a required threshold grade, yet also made sufficient levels of progress. Indeed, there seemed to be a shift away from focusing solely on the grade C threshold for some schools. This shift might have been beneficial to students, or at least certain groups of students, since it may have resulted in resources becoming more evenly distributed, rather than being focused upon those on the grade C borderline. However, it is also another element that schools were required to balance in their decisions, which may have led to other unintended or undesirable consequences.

> *So you're always looking at progress and I think that's become more and more the stick I'm being battered with.*
>
> *(Head of department, Academy F)*
>
> *Yes I think it's one of the things OFSTED are looking for now, they're looking for 4 levels of progress for the majority of students.*
>
> *(Teacher, Academy E)*

In summary, for the majority of schools included in this research, the sense of pressure to achieve results was overwhelming, regardless of whether this was for the school or the students' needs. Furthermore, this appeared to be dictating school's behaviour, in terms of the way that students were taught, prepared and entered to assessments, issues that the remaining themes now turn to.

Approaches

Who, what, when and where?

The focus on results, and primarily the grade C boundary, resulted in schools adopting a highly strategic approach to entering their students to assessments. It was clear from discussions that, although different individuals within the schools were responsible for entries and that approaches varied significantly between schools, considerable time, thought and effort was invested by each school in devising the approach that they

perceived would maximise results. This aligns with previous research that describes the 'interrogation' of school life designed to maximise results (see Ball, Maguire, Braun, Perryman and Hoskins, 2012; Gillborn and Youdell, 2000). While there has been some indication of such practices being used by schools through data analysed for awarding purposes (e.g. see Stockford, 2012a; 2012b), the degree to which schools were being strategic, and the multitude of ways in which this occurred, was quite striking.

This permeated through schools' entry approaches in multiple ways and on several levels. Schools considered *which* students to enter, *when* to enter students, *which* specification and awarding body to enter students with, and at *which* tier. Interestingly though, practices varied, so although schools were typically seeking the same end point – to maximise their results, typically at grade C – the way in which they sought to achieve this differed.

The decisions concerning each aspect of schools' entry approaches were tightly interwoven with one another, meaning that schools had highly individualistic and complex approaches, which also varied for different groups of students. Thus, while there were commonalities in practice, categorising schools according to the way that they entered students to assessments proved challenging (as did coding the data in this manner), due to the diverse range of approaches that were employed. Thus, this theme focuses on describing the lower level aspects of schools' entry approaches, rather than the overall approaches that each school used. To do the latter was considered to risk over-simplifying the complexities identified.

The first key finding was that the majority of respondents reported having used some form of EMEE in the past for some or all of their students. One exception to this was the selective school that, under less pressure to achieve results due to the high calibre of their

students, reported entering their students once at the end of Year 11.[75] This approach was used to support what they perceived to be effective teaching and learning for their students. For the other schools, early entry was typically used to facilitate multiple (or repeated) entry across different examination series, often with different examination boards. The use of multiple examination boards was not unusual and did not seem problematic to respondents: indeed, some reported that their students had been entered with each of the three largest awarding bodies within one academic year. The use of multiple examination boards was also reported in a large scale quantitative analysis published by the DfE (2013b; see Chapter 2 for a summary), that also showed that the practice of entering multiple specifications was more common in mathematics than English. This is perhaps not surprising given that there was no internal assessment in mathematics (since 2009) and the content across specifications is broadly comparable, thus facilitating entry to different specifications.

> *… they will have sat in November with AQA, they sat in January with OCR, they sat in March with Edexcel and then they sat with AQA in June …*
>
> *(Teacher, Comprehensive B)*

Alongside the use of multiple entry across examination boards, there was also evidence of students entering multiple specifications in the same examination series, an approach identified in the literature (see Chapter 2). Again, this is likely to be more common in mathematics since multiple specifications were available (until November 2013 linear, modular and linked pair pilot[76] specifications were available). Thus, students could enter

[75] Note that had the changes to performance measures not been announced in September 2013 then this school had intended to enter their borderline grade C Year 11 students in November 2013 to provide an indication of progress, following the loss of modular examinations.

[76] The linked pair pilot consisted of two GCSEs in mathematics: applications and methods. Until November 2013 this was available in a modular format, following which it was linear. The linked pair specifications were last offered in November 2016.

two specifications in the same series, either with the same or different awarding bodies. While this approach increased the number of opportunities for students to enter, concerns were raised about the potential effects on the students.

> *… we did have a couple, not very many, but we did have a couple who we gave a shot of the Unit 3 and the linear at the same time you know, that was a tough ask for those particular kids but that was another strategy because some of them were right on the edge of achieving for the unitised method and those, you know, it just gives them another opportunity.*
>
> *(Head of department, Academy F)*
>
> *We also use the system where we would enter some students for the Linked Pair pilot as well as their module exam so we ran the two things at once.*
>
> *(Teacher, Academy B)*

The frequent use of EMEE by schools was in no respects coincidental: it was highly strategic and perceived as a means of allowing students, particularly grade C/D borderline students, as many entry attempts as possible, resulting in many students having several attempts at GCSE mathematics throughout Years 10 and 11. Thus, entering GCSE mathematics four or more times was not uncommon.

> *… actually the Year 11s that left at the end of last year, there's some kids in that group actually did the exam in the summer of Year 10, the November of 11, the March of 11 (as there was a March exam period then) and then the summer of 11 as well. So they had actually four shots at the exam.*
>
> *(Teacher, Academy A)*
>
> *So, if I go back to last year, our Year 11s, the vast majority of the students were entered five times across the course of Year 10 and 11. So they were basically entered at every point they could have been right from the start of Year 10.*
>
> *(Head of department, Academy D)*

In addition to considering when students were entered, schools adopted a strategic approach to choosing which examinations to enter and at which tier. Again, this aimed to maximise results. For example, some schools reported switching students between specifications part-way through the course, typically moving their lower achieving students from a modular to a linear qualification if they had performed poorly on the early modules and could not, or were unlikely to, achieve their target grade. Essentially, this 'wiped the slate clean', providing students with another opportunity to enter without having 'banked'[77] marks on their early modules (see also Vidal Rodeiro, 2014).

> *Yes, tried that route as well. I had pupils doing the modular course and if they weren't doing very well in their modules and it looked as though the modular route might lead to a lower grade then they've been entered or switched to enter for the linear model.*
>
> *(Teacher, Academy G)*
>
> *… only people that were very close to the C on unitised were kept on the unitised so most people went down the linear route.*
>
> *(Teacher, Academy E)*

Similarly, determining a student's tier of entry was often a strategic decision, although again, practices varied. Some schools opted to use the higher tier for their borderline grade C students, sometimes because they considered it 'easier' to achieve a grade C on this tier due to the perception that students only needed to achieve a low number of marks. This often led to weaker students being 'trained' specifically for the assessment, an issue that is discussed later.

[77] 'Banked' marks refer to those achieved in previous examination series in a modular qualification. They are banked in the sense that they have already been awarded to students and therefore cannot change.

> *The cohort that go for the multiple entries have always been higher so, even if you've got a pupil on a lower level and they've been identified, they will be trained to do the higher paper because it is deemed as easier to get a C grade on than a foundation.*
>
> *(Teacher, Comprehensive A)*
>
> *If you think the kid's got an inkling of a chance of getting a C, you put them in for the higher level, because they've got less marks to score.*
>
> *(Teacher, Academy A)*

Other schools considered their use of tiering as a means of supporting progression for their middle ability students. Thus, borderline students were entered for the foundation tier initially and then moved to the higher tier to build on their grade. This was considered beneficial for the students since without EMEE, these students would have remained on the foundation tier until the end of Year 11, meaning that their grade would have been capped at a C.

> *… if we hadn't entered them at Year 10, we would have sat on them, they would have done Foundation at the end of Year 11 if we'd not done any early entry. So we entered the C/D borderline cohort in June in Year 10 … those who have passed have now progressed through onto Higher and with the intention of progressing on to getting a grade B.*
>
> *(Teacher, Comprehensive B)*

This practice was reportedly driven by the students' needs and the desire to support progression. However, one might argue that if these students were capable of sitting the higher tier, then they should have been allowed to focus here initially. Otherwise, this practice could result in students only studying the higher tier material for a limited period, which would likely have implications for students wanting to progress to advanced level mathematics.

> *I mean one thing I have noticed with the early entries is that … for instance my set 2 this year, they won't have done 2 years of Higher, they will only have done 1 year of Higher. Alright when you're in Foundation, you're in D and C but you're not doing that good stuff that you need for A Level. So when they get into wanting to do A Level maths, they might have a grade A at GCSE let's say, but are their algebra skills really up to scratch because they've only been doing Higher for 1 year whereas the top sets have been doing Higher throughout.*
>
> *(Teacher, Academy E)*

Tiering and EMEE were also used in a strategic manner for the more able students, again with the aim of securing a grade C. Here, schools reported entering students to the foundation tier initially to 'bank' a grade C, before moving them to the higher tier. This approach did not always sit well with respondents and had clear implications for progression and student motivation, issues that are discussed later.

> *Set 2 are students who are capable of doing Higher but we've been starting them off on Foundation, kind of banking a grade C and looking to progress from there.*
>
> *(Teacher, Academy E)*
>
> *As I say, the school that I trained at was very much … the top end entered Foundation in Year 9 to get a C which always seemed really bizarre to me because I was like they're the kids that you can pretty much bank will get a C or above no matter what …*
>
> *(Teacher, Academy H)*

The manner in which tiering was used to support progression or bank grades is likely to be symptomatic of EMEE. In a system of single examination entry, Gillborn and Youdell (2000) discussed the strategic use of tiering, yet here teachers reported 'playing it safe' and only entering students to the higher tier if they were likely to achieve well above the lowest grade available. Thus, EMEE can facilitate movement between tiers in a way that is not possible with single examination entry.

Standalone early entry

Thus far, EMEE has largely been portrayed as a strategic approach used by schools to maximise grades, primarily for grade C/D borderline students. However, this was not always the case, and there was also evidence of early entry being used as a standalone approach. For example, some schools reported entering their highly able students early with the intention that they would sit another qualification the following year, such as a level 2 certificate in further mathematics. This aimed to bridge the gap between GCSE and A level for those students who intended to continue studying mathematics in post-compulsory education, while also ensuring that they achieved highly at GCSE, and was largely considered an effective approach. However, achieving a high grade at GCSE when entering early was not always a given, even for the highly able. Hence, there appeared to be a fine line between deciding who should, or should not, be accelerated in this manner.

> *… if you got a lot of kids through in November with A and A*s, you could then start this 'bridging' course which was very useful to set them up for A Level because then they're not quite as intimidated when they get to A Level and they're also mathematically much more able.*
>
> *(Head of department, Academy C)*
>
> *With the top set kids who we may have been putting in, a lot of them got their A*s early so we were able to do a further maths qualification … which they really enjoyed. Some of them weren't able enough to do it so, again, you had kids in there who were getting put in for this exam who weren't probably clever enough for and bombing out and that and having sort of a negative view of maths from that experience.*
>
> *(Teacher, Comprehensive C)*

Early entry was also frequently used for what respondents described as 'at-risk' students – those students who were unlikely to remain in education or be available for examinations at the end of Year 11 – with the aim of ensuring that they achieved a grade before leaving school. This was likely to continue, despite the changes to accountability

measures, with respondents considering this to be in the best interests of those particular students.

> *… the other things that we'd use early entry for, and we would still use early entry for, is what we call 'kids at risk'. In other words, we don't think they're going to last until the end of Year 11.*
>
> *(Teacher, Academy A)*
>
> *… we will look at the 'at risk' students and they might be entered at the end of Year 10 … we'll look and we'll say 'right, he probably won't make it, she probably won't make it, we'll sit them and just take the hit on the results.*
>
> *(Teacher, Comprehensive A)*

Flexible and experimental

In addition to schools being highly individualistic in their entry choices, differing both between and within schools for various groups of students, it was apparent that schools' approaches were not static; rather, they changed and developed over time. Thus, schools were often experimental in their approaches, attempting to devise what they perceived to be the most effective formula for achieving results, be it for the school, or for the students.

> *Yes, we adapted … when we entered them, how, how we prepared them, what resources we put in place, all of that was on-going and they were … it was strategic decisions, it wasn't sort of like 'let's have a go with this scattergun on these and these'. It felt like that a bit at the start, it was a trial, it was an experiment, but then it became far more sort of set out and planned and delivered, yes.*
>
> *(Head of department, Academy J)*

This experimental approach often entailed trialling approaches with a particular group of students, before extending it to a larger cohort. In addition, approaches were constantly reviewed.

> *… we opted to put a small cohort in for the linear maths in addition to them still following modular to see how effective it would be for certain types of students. We found that the results were quite promising and so last year we actually extended that to the whole cohort.*
>
> *(Examinations officer, Academy K)*
>
> *… last year we did, sort of, a little pilot where our set 4, we put a lot of them in for the Higher paper and what happened was, they didn't, sort of, achieve a grade C but it did help them to achieve their target grade so they made the levels of progress. So we are thinking about that strategy this year but being very selective as we choose which students to put in.*
>
> *(Teacher, Academy B)*

Schools' approaches to entering students to assessments were developed in response to a number of factors including the availability of assessments (see Chapter 1), the success or otherwise of previous approaches (as judged by examination outcomes; see Gillborn and Youdell, 2000), and external pressures such as accountability measures. Thus, many schools reported doing what they could within the confines of the examination system.

> *… then it's like 'right, well we'll try that' and then you do it once and you're slightly successful and then the head of department, not the head of department, deputy head's like 'well we'll try it again, let's do it again next year'.*
>
> *(Teacher, Comprehensive A)*

The impact of schools' approaches on students was also considered by the majority of schools, albeit to a greater or lesser extent. Thus, a minority of schools reported that they had learnt from what they considered to be 'past mistakes'. This typically centred upon the effects on student motivation, which are discussed later, and primarily related to the higher achievers, where the pressures to achieve a grade C were less, rather than those around the C/D borderline.

> *I think one of the things that I have learnt though is that the top sets, they really don't need to be anywhere near foundation, and I think that was a mistake that we've done over the last couple of years …*
>
> *(Head of department, Academy F)*

Schools' approaches were often highly data driven. This was facilitated by the use of multiple entry, since schools had considerable data on each student's performance following each examination series, information that could feed into future teaching approaches. The level of data analysis undertaken by some schools was quite striking; considerable time was invested in tracking students throughout the GCSE years and targeting interventions where necessary, often using early attempts as formative assessment to enhance their students' chances in future assessments. To some extent, this offered a degree of personalised learning, since teachers could target the areas where individual students had performed poorly. Thus, the loss of such data following the return to single examination entry post-2013 was a considerable concern for some respondents, since this was likely to result in increased workloads and fewer opportunities to support students in this manner.

> *We used to do a lot of analysis of the results from the early entries and then we'd hit key areas that groups were weak on and we haven't got the opportunity to do that now so it means that we have to do that internally which is a massive admin task to mark it and uploading the results for every single pupil on a question by question basis.*
>
> *(Head of department, Academy D)*

Fluid teaching groups

The flexibility of schools' entry approaches extended to the way in which students were taught and, specifically, the way in which students were set into ability groups. Many schools reported placing students into ability groups that were targeted at achieving a specific grade, then moving students between these groups following each examination

series depending on the grade that they had achieved, an approach only really made possible by EMEE. Thus, teaching groups were constantly in flux.

> *It was groups of students but those groups were fluid so, with our internal assessments, we would move quite regularly, probably once every month, you'd get a few movements between classes of different students.*
>
> *(Head of department, Academy D)*

This approach to setting students was primarily used to allow schools to focus on certain students, typically the C/D borderline students.

> *It allowed you to focus on a particular group of pupils more for an intense period of time and then get them the grades they needed and then move on to another group.*
>
> *(Head of department, Academy J)*
>
> *It was also driven by the fact that we were able to shuffle the groups as well and move pupils out of particular focus groups to get groups smaller. It's something we've done a lot of here and I think it's one of the main reasons that we've been able to achieve such a rapid turnaround of results in 3 years' time really.*
>
> *(Head of department, Academy D)*

The group of borderline grade C/D students was constantly in flux throughout the year. Students who achieved a grade C were moved to another group targeted at achieving a higher grade, allowing lower achieving students to join the grade C group. This approach was considered to boost the chances of students achieving a grade C, since schools could target resources and interventions effectively, potentially to the detriment of other students elsewhere. In addition, the size of this group of students tended to decrease throughout the year (facilitated by EMEE), allowing more attention to be given to individual students. This could be considered an example of the 'exam factory' approach that has been discussed previously in this chapter.

> *You'd enter your students in Year 10, you'd find out how many have got Cs, you'd put them in sets that … so you'd almost say they're 'banked' scores and then the ones that were left, you'd then focus heavily on. So intervention strategies, put them in the same groups, do all these kinds of booster sessions … pull them out of other lessons …*
>
> *(Head of department, Academy C)*
>
> *So, as soon as they got the C, they were out and then basically the gaps would be filled with people who were still on a D after the previous exam period.*
>
> *(Teacher, Academy A)*

In some respects, this approach could be motivational, since students could be motivated by the success of their peers.

> *… they looked around and thought 'hold on here I'm still in the group', you know, 'people are coming in, people are leaving but I'm still here', you know, and it did have a positive motivational effect on them.*
>
> *(Teacher, Academy A)*

However, there were also negative consequences, since students had little stability in terms of who they were taught by.

> *I know that when students have changed groups and they've come back and they've said 'it's different' and of course it's going to be different because everybody teaches differently with different styles and that isn't a good thing because you need continuity …*
>
> *(Teacher, Comprehensive B)*
>
> *You also never really built those kind of relationships which is essentially what teaching is about.*
>
> *(Head of department, Academy J)*

Furthermore, this approach could also cause disquiet amongst teaching staff, since it could suggest that some teachers, specifically those teaching the borderline students, were more proficient than others.

> *… it created a bit of ill feeling amongst staff and rightly so because it created a bit of a 'you get the better classes' or 'you're supposedly a better teacher than us', that kind of, you know, animosity and unpleasantness.*
>
> *(Head of department, Academy J)*

The movement of students between teaching groups seemed to rely heavily on the notion that material in the assessments was targeted at a particular grade. Given that this assumption has been challenged (e.g. Stobart, Bibby and Goldstein, 2005), the appropriateness of these approaches could be questioned. However, further information on the way in which these groups were taught would be required to evaluate this further.

Exam training

The focus on results and associated entry approaches had clear implications for teaching and learning, both throughout the academic year and around the time of each examination series. Thus, schools could also be considered to be strategic in the way that students were taught and prepared for assessments. This primarily focused upon the use of 'exam training', where interviewees reported training their students to pass examinations, potentially to the detriment of their learning. Typically, this was concentrated on the grade C/D borderline students and was facilitated by the movement of students between teaching groups, as described above. The use of exam training is, to a certain extent, likely to be inevitable, even in an examination system where students are entered only once. However, the use of EMEE is likely to have facilitated some aspects of this and increased the frequency of its use.

> *… they're just trained to pass the exam.*
>
> *(Teacher, Comprehensive A)*

The concept of exam training centred upon the use of 'topic spotting' or 'teaching to the test' for some respondents. There was evidence that teaching was focused on certain

topics that were anticipated to appear on the examination papers, or topics that teachers' perceived their students would be able to score marks on. This meant that some students studied a narrowed curriculum, rather than the whole programme of study. Clearly, this has potential implications for those students who may wish to pursue mathematics in post-compulsory education, as well as those who don't, in terms of ensuring that the necessary mathematical knowledge and skills to succeed in the future are secured. In a sense, teaching was not targeted towards developing long-term mathematical understanding. Rather, the imperative to achieve results at a particular time point took precedence.

> *… you would obviously chuck stuff out that you know has created problems over the years but … them kind of kids, I'm not going to hammer away for a week trying to get them to rearrange formula, it's just not happening ...*
>
> *(Teacher, Academy A)*
>
> *I know my Year 11s that left last year, because we were doing all sorts of different exams, we were just constantly teaching to the test and that is not going to be good for them mathematically.*
>
> *(Teacher, Academy H)*
>
> *… there's no point teaching the A and A* stuff because they wouldn't be able to do it in the exam … So, even though we could have probably … some of the kids had the capabilities to deal with it, in the exam they wouldn't have a clue so we focused on getting the B to C grade stuff topped and tailed and really focused on that.*
>
> *(Teacher, Comprehensive C)*

The use of exam training was primarily driven by the desire for students to achieve a particular grade, typically a grade C. This led to a perception by some respondents that teaching the whole curriculum would be a 'waste of time', since this was not considered the most effective way to ensure that students achieved their target grade.

> *… but I'm not going to waste kids time going through stuff that I know they ain't never going to have … so you use the kid's time hitting all the big stuff, the big scoring stuff, the banker stuff, you know.*
>
> *(Teacher, Academy A)*

> *But also there's no point teaching a topic if you know they're not going to get the marks …*
>
> *(Teacher, Academy E)*

The use of 'topic spotting' was facilitated by the perceived predictability of examination papers. Some teachers reported creating a checklist of the topics on each examination paper, then using this to predict future questions. This, in turn, shaped their teaching and the topics that they focused on.

> *I used to say to the kids, 'look, I'm not saying this might be on the exam, I'm telling you it's going to be on the exam.*
>
> *(Teacher, Academy A)*
>
> *It was the ones that we knew were very likely to come up and historically the ones our students pick up a lot of marks on so cumulative frequency diagrams for example.*
>
> *(Head of department, Academy C)*

In addition to 'topic spotting' and 'teaching to the test', exam training also entailed repeated periods of revision prior to each examination series. For those who entered students repeatedly throughout the academic year, this created a culture of constant revision, resulting in less time being available for everyday classroom teaching.

> *Well I think the key thing to say is that, because of multiple entries, the maths curriculum for Foundation pupils in Years 10 and 11… has been a continuous cycle of just being prepared for an exam.*
>
> *(Teacher, Academy G)*
>
> *… at the end of Year 10, throughout Year 11, became an exercise of just revision constantly because you were just preparing for the next exam that was only a couple of months away each time.*
>
> *(Teacher, Comprehensive C)*

There was an obvious tension for many interviewees between the use of exam training and the negative effects on teaching and learning, yet the drive to achieve results, either for the school or the student, appeared to take priority.

> *It's not really educating them, it's training them to sit an exam so that didn't sit well with me but it was … at that school especially, it was a numbers game.*
>
> *(Teacher, Comprehensive C)*

This tension was highlighted by the manner in which many respondents welcomed the changes to performance measures announced in September 2013. It was anticipated that, following these changes, students would only be entered once at the end of Year 11, bringing an end to the culture of constant revision and examinations. This was seen to facilitate a more in-depth approach to teaching mathematics, although it was acknowledged that some form of exam training, or 'teaching to the test', would still persist.

> *In one way I feel this system is better because we're going longer periods of time and it helps us to concentrate and sort of have a richer maths experience because obviously we can concentrate on what the students can understand and not understand rather than 'these are the exam questions, we've got to be able to answer this sort of thing', so it's a different emphasis …*
>
> *(Teacher, Academy B)*
>
> *Yes we'll still teach to the test, yes, because the test is so important … It's still the same exam, we're just only getting one go at it so all those techniques we've used before will still be relevant in this one. It just won't be so heavily focused I think at different parts of the year.*
>
> *(Head of department, Academy C)*

In summary, the desire to maximise results has seemingly resulted in a number of approaches being adopted by schools, relating to how and when students are entered for assessments, and how they are taught. The following section discusses the potential implications of these approaches for schools, students, and the wider assessment system.

Implications

Rising outcomes, but poorer mathematicians?

The use of EMEE and exam training, as described above, was perceived to have implications for students' outcomes, both in terms of the grade that they achieved and the knowledge and skills that they acquired. In terms of grades, there was a perception that EMEE would lead to outcomes increasing at the national level, particularly at the grade C threshold.

> *I think the national improvement in A* to C is probably a large amount to do with repeated entries and I think a lot of schools are probably doing what we're doing and shuffling pupils afterwards.*
>
> *(Head of department, Academy D)*
>
> *I think the percentage of Cs must be creeping up. If people are having 4 goes as opposed to 2, you know, I think they must be creeping up …*
>
> *(Teacher, Academy E)*
>
> *I think it must have positively influenced C and above certainly. Maybe not so much A and A*, it might have pulled it down if anything, but certainly the threshold targets I would have thought it, sort of … naturally inflated them.*
>
> *(Head of department, Academy F)*

However, this rise in outcomes was not necessarily considered to be solely related to, or indeed representative of, the mathematical ability of the cohorts. Rather, outcomes were perceived to be artificially inflated due to EMEE and associated teaching practices.

> *I think it provides a false confidence nationally in terms of maths education and you don't really get a true picture of the maths ability in the country.*
>
> *(Teacher, Comprehensive C)*

> *I think because of the game playing with early entry I do think that's increased the results. I don't think that's a true reflection on the ability of the nation … so the fact that the grades have gone up, if they had, I wouldn't say that was a reflection on that the kids are getting brighter, it's that we're getting better at ticking boxes.*
>
> *(Teacher, Academy H)*
>
> *I think at the moment what's happened is we've got this kind of artificially inflated grades … I think it's because there's this pressure on the league tables and those results you have to report that schools do as much as they possibly can but it's artificially inflating everybody's grades.*
>
> *(Teacher, Selective A)*

The perception that EMEE would impact on outcomes is further evidenced by the way that respondents reacted to the changes to accountability measures announced in September 2013, with many participants suggesting that a single examination entry would depress results.

> *I think you will get a slight slump in results because people don't … can't have that security, you know, with intervention and what not.*
>
> *(Head of department, Academy C)*
>
> *I'm really interested to see whether there's a dip this year. I think there's a good chance there could be a slight dip as a result of removing it on the headline A* to C figure.*
>
> *(Head of department, Academy D)*

Nonetheless, these changes were largely welcomed by many respondents, since they were considered to promote a better mathematics education for students. Further, single examination entry was considered to level the playing field and provide a better indication of a student's ability, a view shared by interviewees from contrasting schools, as illustrated below.

> *... it's nice because it's a level playing field. I was always slightly uncomfortable that you'd got pupils who were having multiple entries on exams and ours were always going in at the end and it seemed a bit unfair.*
>
> *(Teacher, Selective A) [High Achieving school]*
>
> *… it gives a purer result I think of your students' attainment. So, from that respect, it's quite good … everyone's in the same boat.*
>
> *(Head of department, Academy C) [Low achieving school]*

Despite this, single examination entry was considered to penalise certain students, particularly those from disadvantaged or difficult backgrounds.

> *… an outstanding school where you've got, you know, grammar schools, independent schools, where you've got a massive cohort of very intelligent kids with very supportive parents, they probably could enter a year early and know that they'll get As and A*s whereas here we can't so it penalises the lower, well the lower class. I'm going to say lower class; I'm going to use those words.*
>
> *(Teacher, Comprehensive A)*
>
> *What it does I think is it'll further separate grammar schools and high achieving leafy schools to inner city comprehensives where you get children who don't attend after the Christmas of Year 11 and things.*
>
> *(Head of department, Academy J)*

The perceived disconnect between rising outcomes and mathematical ability was considered to be due to teachers becoming more successful at training students to pass examinations. Furthermore, schools were considered to be 'spoon-feeding' students in order for them to achieve results, potentially to the detriment of their learning.

> *… I think it has improved results and I think the reason why is because the maths teachers have become much, much better at preparing pupils to be successful in exams …*
>
> *(Teacher, Academy G)*

> *I was very good at training children, you know, 'do this, do this, it doesn't matter why, do it' and you could train them to do that very quickly but mathematical understanding? No.*
>
> *(Head of department, Academy J)*
>
> *… the week before an exam, all of my classes would be covered and I'd have Year 11 all day and we'd have a full day on … so it was basically sending a message to the kids, 'don't revise at home because we'll change your timetable, we'll take you out of your lessons and we'll revise with you then', so it was really spoon-feeding them and that then created a culture of, you know, 'I don't need to do any work at home, I don't need to do anything by myself, everything will be done for me'.*
>
> *(Teacher, Comprehensive C)*

Thus, the use of exam training was perceived to have negative implications for what students know and can do – and even their approach to studying – regardless, to a certain extent, of the grade that they achieved. For example, many respondents considered there to be a difference between students that had been trained to pass their GCSE and had entered on multiple occasions, and those that had not, even when the same grade or UMS scores were achieved.

> *… if you've got a student who's got an A grade from this school and a student that's got an A grade from an external school like a high school in XXX … when you teach the Year 12 maths, you generally find the pupils in this school are much, much stronger and the pupils from the other schools who … and you look at their UMS scores, they're comparable, they're significantly weaker because they have been … they multiple entry and they've been coached and coached and coached and hammered far more than they do here.*
>
> *(Teacher, Selective A)*

This was considered likely to impact upon how successful students would be if they continued studying mathematics in the future, or if they required mathematics in a future job role.

> *Well we've got students here who've come to do A Level from other schools who have got As and you think 'oh A, very good' but they're struggling with it because they were just coached constantly to get that A grade. They're not necessarily particularly strong mathematic ability-wise …*
>
> *(Teacher, Comprehensive C)*
>
> *… through the exam training, they're coming out with C grades and then you, sort of, down the grapevine I hear from, like, colleagues and things who work in sixth forms, you get a child with a B grade at GCSE, they should be at a good enough level to be able to do A Level maths and they don't even know how to add a fraction …*
>
> *(Teacher, Comprehensive A)*

Further, it raises questions regarding the extent to which the grade that a student achieves can, or indeed should, be taken at face value, both when allowing students access to further education or when choosing potential employees, since the relationship between a student's mathematical ability and the grade that they attain could be mediated by the way in which that grade was achieved (see Chapters 6, 7 and 8 for further discussion).

> *… like I look at last year's cohort and you've got students in the top 2 sets, like set 1, leaving with a C grade and then you've got pupils who used to be in set 5 leaving with a C grade. Now you sit those 2 people next to each other there is a massive difference in their ability levels but yet they've both got C grades and on paper they're both as equally employable whereas this person down here, the only reason they got a C grade is because they were dragged out of loads of lessons, put in intervention classes and trained …*
>
> *(Teacher, Comprehensive A)*

Motivation and engagement

In addition to the impact on students' grades and knowledge, EMEE and associated teaching practices were frequently discussed in relation to the perceived effects on students' motivation and engagement, both in a positive and negative manner. While there was considerable variation, the overwhelming effects were considered by

respondents to be negative. These effects were not restricted to particular groups of students and there were considerable differences between schools, suggesting that the relationship between EMEE, teaching approaches and motivation is complex. Furthermore, it is likely to be mediated by several factors including the attitudes within a particular school and the effectiveness of individual teachers in engaging students.

Primarily, respondents discussed the negative effects of EMEE in terms of a loss of engagement and motivation once a student had achieved their target grade, whatever this grade was, and whenever this occurred. Thus, students were thought to be disinclined to continue studying mathematics to try to improve their grade, even if they were considered to have the ability to do so. This resulted in students becoming disengaged and at times, disruptive to their peers.

> *So it was basically this race for mediocrity rather than pushing those kids who, some of them, could have definitely got A*s, definitely, but because they had got their target grade, they sort of threw the towel in and that happened from Grade A kids, you know, down to Grade C.*
>
> *(Teacher, Comprehensive C)*

This loss of motivation and engagement was considered to be particularly acute once students had achieved a grade C. This attitude is perhaps not surprising, given that students were reportedly well aware of the importance of achieving a grade C. Thus, some students were reported to 'switch-off' once their grade C was 'banked', since this was the grade that they needed to access further education or employment and there seemed little reason to improve this.

> *Another problem we've had is that they do the college interviews now and they get the conditional offer of, you know, you've got a place if you get a C, well some pupils already have a C so they go 'oh I'm already into college so what's the point?' and then they give up and then they don't get the Bs and the As …*
>
> *(Teacher, Comprehensive A)*

> *As soon as someone got the Grade C where that was their target they took their foot off the gas. So you had whole groups of kids who were disengaged.*
>
> *(Teacher, Comprehensive C)*

Motivation issues were also thought to affect the higher ability students, particularly those who entered on multiple occasions or who entered the foundation tier initially to 'bank' a grade C, before moving to the higher tier. This approach was subsequently rejected by some respondents, due to the perceived negative effects on students.

> *The A – A* pupils as a group were losing motivation as well. I think, you know, the quantity … the five times that they were entered wasn't necessarily the right decision on our part to be honest with you …*
>
> *(Head of department, Academy D)*
>
> *We have in the past … our top set of students were made to do Foundation at one stage. I think a couple of years ago they were made to do Foundation in November against the advice of myself and my second in department at that time …*
>
> *(Teacher, Comprehensive B)*

This suggests that increased EMEE and associated teaching practices may have resulted in increasing numbers of students becoming disengaged with studying mathematics throughout their GCSE years. Hence, although most schools required the majority of their students to continue studying mathematics until the end of Year 11, contrary to speculation and previous evidence (see Chapter 2), the quality of learning for certain groups of students during this period was perhaps questionable.

> *All the way through Year 11 we stuck with maths, 3 hours maths, we didn't switch to statistics or anything like that.*
>
> *(Teacher, Academy E)*

The loss of motivation was considered a particular issue for the students that were in Year 11 when these interviews were conducted, many of whom had expected to sit an examination in November, but were withdrawn following the accountability changes announced in September 2013. The majority of schools reported experiencing considerable difficulties in motivating these students.

> *… this year's Year 11 seem to have had a rough deal to me because you've got this, now they're treading water for 4, 5, 6 months, and that is so easy for kids to switch off ...*
>
> *(Teacher, Academy A)*
>
> *… I've now finished the syllabus before Christmas, I've got 4½ months of time to kill and these are pupils that are not motivated by maths and it would have been nice to get them get their C or whatever, C or B, and then we could think of other strategies, things to do with them for 4 or 5 months because I feel they've peaked now and this next few months are they going to continue to improve or will they actually tail off?*
>
> *(Teacher, Selective A)*

Although the discussion above suggests that the effect of EMEE on students' motivation and engagement were considered to be negative, this was not always the case. Some respondents reported that, for some students, EMEE appeared to have a motivational effect. Mainly, this was for those students that could build on their previous result in a later examination series. For example, students who were close to achieving a higher grade in one examination series could be motivated to strive for this in the next examination series.

> *… some of the people were quite motivated as well when they saw how close to a C they'd been or how close to an A they'd have been in the top set …*
>
> *(Head of department, Academy D)*

Weaker students, in particular, were considered to benefit from entering early and achieving a grade then being able to build on this. Often, this was due to these students lacking confidence initially.

> *Those 11 pupils … as they sat it across 5 times, they saw themselves go from Us to start getting grades to going 'oh right …', and then … when they were getting Ds, you really saw … they weren't difficult lads anymore. They were a team of lads who went 'right, we're going to get Cs'.*
>
> *(Head of department, Academy D)*
>
> *I took them over at the start of Year 10 and they were a little bit low on confidence and so I focused in on the D and C topics and … every one of them passed with flying colours on Foundation, with the exception of one of them who doesn't attend, and the confidence has grown and … we're focusing on the Higher.*
>
> *(Teacher, Academy E)*
>
> *If you take the less able students, as I mentioned earlier, it was a way for them to actually see … they might know within themselves that they're not going to get a Grade C or a B because it's a long way but if they see themselves getting an F and then an E, it really does encourage them. It makes them think 'actually I might get the D; I could do this' and so it kind of encourages them to go on and it gives them a very positive view.*
>
> *(Examinations officer, Academy K)*

Realising potential

One group of students that were considered to be particularly negatively affected by EMEE and associated teaching practices, and who had significant issues in terms of motivation, were what could be termed 'the forgotten middle': the group of students who, in terms of ability, were positioned between the borderline grade C students (where schools focused most of their attention) and the highest achievers (who were likely to achieve highly regardless of when they entered).

The primary concern centred upon how EMEE might limit, or reduce, the opportunity for these students to reach their potential.[78] Many respondents felt that EMEE to maximise the chance of achieving a grade C was having a negative effect on these students who

[78] Note that students' potential was considered in terms of the grade that a student achieved, not the level of mathematical understanding that they reached.

were considered to be capable of achieving a higher grade. There are two principal mechanisms through which this was thought to occur: first, schools tended to focus resources on the C/D borderline students, potentially to the detriment of other students; and second, students tended to lose motivation once they had achieved a grade C, as discussed previously in this chapter.

> *I think there was perhaps 3 individuals that I think probably would have been able to get a B that didn't just because of motivation issues that were mentioned earlier and, some of our A/A* students, I think we may have had 2 individuals that could have had an A* had they just sat one summer exam but achieved an A.*
>
> *(Head of department, Academy D)*

In many respects, schools appeared to treat this 'middle' group of students in the same manner as those who were on the C/D borderline when, in reality, these students were likely to achieve at least a grade C regardless of whether they entered multiple times or not. This might be due to a lack of provision or consideration about how these students should be treated, since attention was often focused elsewhere. Essentially, these students could sometimes be considered as 'forgotten'. For example, in some instances students were not given the support to improve their grade once they had achieved a C, so while they sat the assessments again at the end of Year 11, they were not taught any more mathematics leading up to this.

> *Teacher: That has happened actually in previous years; they've been removed from maths lessons and weren't given the option to go for like a B grade or better.*
>
> *Examinations Officer: Oh no, they … or maybe can but they've just not had the lesson time to probably progress.*
>
> *Teacher: Yes so they've been entered for the exam but they've not been given any lessons …*
>
> *(Teacher and Examinations Officer, Comprehensive A)*

Alternatively, students' attention could be refocused elsewhere and they were allowed to study other subjects, such as English, during the periods that were assigned to mathematics. This could arguably be beneficial to the students, since achieving a threshold grade in a subject such as English was also highly desirable. However, the logistics of moving students between lessons often prevented this becoming a viable approach.

> *Yes, I mean, initially, a few years ago, a few years ago, there was a lot of flexibility with regard to, you know, working between departments and if a kid had got what they needed in maths but they were still off in English then they would go up to English and do some work. That seemed to last for one summer really and then people started grizzling about kids moving all around the place.*
>
> *(Teacher, Academy A)*

The notion that EMEE might result in certain groups of students not realising their potential and making sufficient progress was evidently becoming a concern for schools, particularly since progress was being used as a measure by Ofsted (see Chapter 1). Thus, in addition to securing grade Cs for performance measures, schools were increasingly aiming to ensure that students made the expected level of progress between KS2 and KS4. While these aims may seem aligned with one another – achieving as much progress as possible should be coupled with ticking the grade C box (for middle or high achievers at least) – this relationship appeared to be complicated by the desire for schools to 'bank' C grades early to reduce some of the pressures at the end of Year 11, and the subsequent lack of motivation for students to improve their grades. However, the desire to achieve a certain level of progress, alongside achieving a particular grade, is likely to support learning and achievement for certain groups of students.

> *We can't really put a set 2 in at the end of Year 10 to bank a C because it would have a massive impact on our league tables with the 4 levels of progress.*
>
> *(Teacher, Academy E)*

> *Levels of progress is the big one now. It was getting them Cs and getting the A/A* group, it is now, we've nailed that, that's fine, but levels of progress-wise, we'll still behind and we need to get that up.*
>
> *(Teacher, Academy H)*

Disquiet amongst teachers

Thus far, this chapter has identified how the pressure to achieve results has driven behaviour in schools, prompting widespread EMEE and associated teaching practices. This raises questions of how the situation described above impacts on teachers, and where responsibility for these events should lie, issues that the final two themes discuss. While both of these issues emerged as themes from the data, they are considered to sit outside the framework used to organise the preceding themes (shown in Figure 13), since both issues interact with and permeate through the themes discussed above.

Turning first to the impact on teachers, it was clear from discussions that EMEE within schools did not sit well with many respondents. While this may have been alluded to throughout this chapter and illustrated by certain quotes, it is worth making this explicit here since it would be remiss to suggest that EMEE did not cause concern for respondents. Rather, it was clear that, although teachers were adopting approaches to maximise results, this was frequently coupled with an overwhelming sense of frustration and disquiet. This related to several aspects of schools' behaviour that have been discussed throughout this chapter. For example, one respondent reported their reluctance to teach the higher tier to weaker students, despite this being considered an effective method of facilitating achievement at grade C. This reflects previous findings concerning entries to tiered assessments, where teachers felt forced to act in a way that was not necessarily in the best interests of the students (Gillborn and Youdell, 2000).

> *I am very reluctant to teach the Higher in the way that I'm just teaching this, this, this and this to make sure I get my 25 marks or whatever it is to secure a C. I don't think, as a Mathematics teacher and an educator, that's scandalous to be honest.*
>
> *(Head of department, Academy F)*

Other respondents reported their dissatisfaction with the concept of training students to pass examinations and the high levels of support given to students when revising for assessments.

> *I think you fall into this trap of exam training and I never came into teaching to train a kid to pass an exam, I came into it to teach maths and it does become training and it's not what you're at school to do.*
>
> *(Teacher, Comprehensive A)*
>
> *There was no emphasis on 'right, it's your exam, you've got to revise for it', none of that because they were so terrified that they wouldn't do it and wouldn't be mature enough to do it so we had to hold their hand the whole time and walk them through how to revise and spoon-feed them revision materials and I hated that aspect of it.*
>
> *(Teacher, Comprehensive C)*

In many respects, schools could be considered to be in an almost impossible situation, leading to a moral dilemma for many respondents. On the one hand, respondents considered it their moral duty to get students a grade C, so would use whatever methods necessary to achieve this. However, on the other hand, individuals considered the use of certain entry and teaching practices – enacted for the benefit of the school – as immoral, due to the negative implications for students' learning. These two imperatives seemed to be pulling in opposing directions. This resonates with previous findings – for example, Gillborn and Youdell (2000) highlighted the dissonance between the morally or socially acceptable approach and the approach that will produce the best results when discussing setting students by ability (see also Perryman, Ball, Maguire and Braun, 2011).

> *And I didn't think we'd be doing our job right or teachers certainly didn't feel as though we'd be doing our job right if we weren't giving these people this opportunity.*
>
> *(Head of department, Academy F)*
>
> *I mean it was an issue that was a real bugbear for me when I was at my last school. I didn't really agree with this sort of racing to mediocrity.*
>
> *(Teacher, Comprehensive C)*

Typically though, the imperative to achieve results took precedence, leading schools to feel that they were competing against one another to succeed.

> *You know, schools, at the end of the day, it's one massive competition, we're all competing with each other and I think everybody, and it's a little bit cut-throat, but everybody tries to get ahead of everybody else …*
>
> *(Teacher, Academy E)*

This tension may have been particularly acute for the respondents in these interviews, since they were largely classroom teachers who were interacting with students on a regular basis. Had the views of senior management been gathered, then the responses may have been different.

> *I mean I don't think any of it was good. I didn't enjoy it. I didn't want to do it and I'm glad I'm not having to but I don't know if I was a Head of one of those schools that I would've necessarily done anything any different.*
>
> *(Head of department, Academy J)*

Alongside the disquiet amongst respondents due to the pressure to achieve results, frustration was also expressed by some respondents regarding the constantly changing climate that participants were working in. This was considered to make the jobs of teachers even more difficult. This might have been particularly acute in mathematics, due to the high volume of changes to the specifications that have occurred in recent years

(e.g. the move from three to two tier assessments in 2008, and the removal of coursework in 2009). However, the speed that policy changes can occur in education suggests that other subjects and qualifications are likely to be affected as well.

> *There's more reform and there's been that much change in maths. I mean, you know, your bog standard maths teacher that's been teaching maths for the last 10 years let's say has just seen so many changes, it's just really difficult.*
>
> *(Teacher, Academy E)*
>
> *It was four changes in four years something like that we went through and it's not fair really because as a teacher you've still got to do that, you've got to do what you can to put them in. It's re-writing schemes of work then re-writing them again and it … it wasn't, I mean the exam board partly were getting their hands forced, I know that, but you feel like you're not working together in the best interests of children.*
>
> *(Head of department, Academy J)*

Responsibility

A final but related issue concerns that of responsibility, and where responsibility for the situation described in this chapter should lie. Although this issue was not raised directly during the interviews, it naturally emerged during discussions and permeated many of the other themes.

For many respondents, responsibility was placed firmly on the Government and there was a sense that schools were merely acting as agents within the assessment system to something that was being dictated to them. Thus, interviewees felt that their hands had been forced by the constant pressure in the system to achieve results. Indeed, many individuals openly admitted to 'gaming the system'. The willingness to admit this so freely may be due to a sense that schools were not responsible for their actions; rather, respondents considered that they were forced to adopt particular behaviours due to the pressures of being held accountable.

> *… we've been 'playing the game' for the Government and, actually, our pupils don't have a good knowledge of maths, they have a good knowledge of an exam …*
>
> *(Teacher, Comprehensive A)*
>
> *They made, you know, they made everyone play the game. There was so much pressure on everyone, 'you must get this amount of Cs, you must do this, you must do that', so schools have fitted in with that …*
>
> *(Examinations officer, Comprehensive A)*
>
> *… you know, it's not us … our school didn't make the policy on what was the right grade. If there wasn't a league table or it wasn't published then it might even be a better system but that's not the case so we have to push the Cs as our main driver and we do.*
>
> *(Head of department, Academy C)*

This led many respondents to report that they were simply doing whatever they could within the confines of the examination system.

> *… we've really just responded to whatever the Government have put available and we, like every other school, will just take the best, the optimum route, through what the system allows.*
>
> *(Teacher, Academy B)*
>
> *… everyone was playing the numbers game so, last year, when we entered pupils 5 times, we did that because we could.*
>
> *(Head of department, Academy D)*

For some respondents there was an inevitably about this, and it was acknowledged that, once a measure of success was in operation, this became the goal.

> *Well if you're going to be measured, if somebody says to you 'you're going to be judged on this', everything else goes out the window and you do what you have to to be judged on whatever that is and that's whatever isn't it?*
>
> *(Head of department, Academy J)*

The sense that schools were merely reacting to Government policy was reinforced by the language that respondents used during the interviews, particularly regarding the accountability changes announced in September 2013 (see Chapter 1). Many participants referred to these changes as 'banning' the use of EMEE, when schools would still be free to use whatever approach they wished. However, there could be implications for their results. Thus, schools considered the changes to be penalising certain students.

> *What they've done is there's no flexibility at all; they've really just said no early entry.*
>
> *(Examinations officer, Academy K)*
>
> *I've got a G and T [gifted and talented] boy in Year 9, he could sit his exam at the end of Year 9 and get an A*, he is that intelligent but he now will never get an opportunity to be pushed and that's the argument for early entry and now the Government are penalising schools for that.*
>
> *(Teacher, Comprehensive A)*

Discussion

Entry approaches

In summary, this chapter has described how the pressure to achieve results has driven behaviour in schools, prompting widespread EMEE and associated teaching practices. The entry policies used by schools are highly complex, requiring schools to invest considerable time and effort in developing and refining these. Furthermore, they are highly individualistic, both between and within individual schools, meaning that a number of approaches were identified. However, by far the most popular approach at the time these interviews were conducted was some form of EMEE. This included:

- Early entry for borderline grade C/D students
- Early entry for high ability students (including acceleration or enrichment)
- Early entry for 'at-risk' students

- Multiple entry across examination series
- Foundation tier entry followed by higher tier entry to support progression for middle ability students
- Higher tier entry for weaker students to target a grade C
- Foundation tier entry for high ability students to 'bank' a grade C
- Entry to multiple awarding bodies across examination series
- Entry to multiple specifications across examination series
- Entry to multiple specifications in the same examination series

This list highlights the diverse approaches within EMEE that were identified in these interviews, with most schools reporting that they used a combination of these. Schools also varied their approaches for different groups of students, often depending upon their ability or their target grade. Their approaches also changed over time in response to several factors including the previous success (or otherwise) of their approaches, the availability of assessments, and external pressures such as performance measures.

These findings provide additional insight into schools' approaches to EMEE than have previously been available and also help to explain some of the trends observed in Chapter 4. For example, the analyses in Chapter 4 identified that it was the students with lower prior attainment that increasingly entered early, and that schools with lower performance outcomes were more likely to enter students early. This is not particularly surprising given that the findings from the qualitative phase suggest that the key driver of EMEE is the pressure that schools feel from performance measures – pressures that are likely to be greater for lower performing schools.

The findings from the qualitative phase also chime with the quantitative analyses in other ways. For example, schools reported entering students early to the foundation tier to 'bank' a grade C, which is likely to have contributed to the greater rise in early entry observed for the foundation tier (see Chapter 4). Thus, the qualitative findings are able to

support and explain some of the trends in the quantitative phase, supporting the mixed methods approach to this research.

Preparing students for assessments and implications

In addition to their entry approaches, schools also carefully considered how they prepared their students for assessments when they entered students early and on multiple occasions. Students were often coached to pass examinations with teaching focused upon certain topics that were (considered) likely to appear on the assessments, or those where students were (considered) likely to achieve marks. To facilitate this, schools set students into teaching groups that were targeted at achieving particular grades, frequently moving students between these after each examination series. This fluidity allowed schools to focus resources on particular students, typically those on the C/D borderline. This was considered an effective method of increasing the chances that a student would achieve a grade C. It is worth bearing in mind that such approaches were not necessarily restricted to EMEE, yet the practice of entering students on multiple occasions facilitates much of this – for example, the ability to move students between teaching sets after each examination series.

On a wider scale, EMEE and associated teaching practices were thought to have increased the proportion of students achieving the grade C threshold at the national level. This perceived increase in outcomes was not necessarily considered representative of the mathematical ability of the cohort though. Rather, EMEE and associated teaching practices (such as 'teaching to the test' and the constant cycle of revision and assessment) were thought to have artificially inflated outcomes, prompting concerns regarding the effect on students' learning. Since the possible impact of EMEE on examination outcomes is considered in Chapters 6 and 7, it is not discussed further here.

While these findings might lead one to criticise schools' use of EMEE, it would be remiss not to acknowledge the importance of grades, both for a student's future prospects and for

schools. Essentially, schools are required to balance the needs of the school and the needs of the students – or in other words, to balance the need to achieve a particular grade with the need to provide good teaching and learning opportunities. Unfortunately, but perhaps understandably, the scales seemed to be tipped in favour of the grade at the time that this study was conducted (the same would likely be true more recently), leading schools to adopt a number of approaches to maximise results.

The desire within schools to maximise results is certainly not a new phenomenon and the impact of performance measures on behaviour, particularly teaching and learning, has been well discussed in the literature (see Chapter 2). What the current study adds is evidence concerning the impact of these pressures on how students are entered to assessments, particularly focusing on EMEE, evidence that is less readily available (see Chapter 2). Thus, this chapter furthers our understanding of the drivers behind schools' entry approaches, suggesting a clear link between the pressures of performance measures and EMEE. Furthermore, while EMEE in the format described here effectively ceased following the changes to performance measures announced in September 2013 (see Chapter 1), it is not unreasonable to think that schools will have, and will continue to, explore alternative methods of maximising results while the pressures of performance measures exist. For example, Stobart (2008) describes the '*continuous game of regulatory cat-and-mouse*' that ensues when schools attempt to meet targets while loopholes in the system are closed around them.

Summary of Chapter 5

To summarise, the qualitative study discussed in this chapter has highlighted how the pressure to achieve results has driven schools' approaches to entering students to GCSE assessments, primarily through EMEE. The study has also identified the methods that schools use for teaching and preparing students for assessments in the context of EMEE. These findings provide additional context to the rise in EMEE than has previously been

available in the literature and help to explain some of the trends identified in the quantitative analyses in Chapter 4. The remainder of this thesis considers the implications of EMEE for examination outcomes and examination standards.

Chapter 6 – Multiple entry and the implications for outcomes

Overview

The present and following chapter build on the findings in Chapter 5 by exploring one of the possible implications of EMEE in GCSE mathematics – the implications for examination outcomes. In the current chapter, empirical data is used to consider the extent to which students entered on multiple occasions, the characteristics of students following different entry routes, and the associated effect on examination outcomes. Following this, Chapter 7 uses simulations to model the extent to which multiple entry might impact on examination outcomes as the proportion of students entering on multiple occasions changes. These two sets of analyses, using different quantitative approaches, complement one another and collectively provide greater insight into the research questions addressed in these chapters.

Introduction

This chapter builds on the findings in Chapter 5, where schools' entry approaches to GCSE mathematics, the drivers behind these approaches, and potential implications were explored using qualitative methods. Specifically, this chapter explores one of the potential consequences of EMEE identified in Chapter 5 – the effect on examination outcomes. In this case, examination outcomes refer to the grades that students achieve. These grades can have significant implications for an individual's future life chances and for school performance measures (see Chapters 1 and 5), yet there are also implications when grades are aggregated across a cohort of students and reported at the national level. For example, national examination results are reported each August and are subject to considerable media scrutiny, with small changes in outcomes prompting debate around whether standards have improved or declined (see Chapter 1).

The potential effect of EMEE on national examination outcomes was discussed by several respondents in the qualitative study (see Chapter 5). Here, respondents considered that frequent EMEE would increase outcomes at the national level, particularly around the grade C threshold, due to teachers becoming more effective at training students to be successful in examinations. This perceived rise in outcomes was not necessarily considered to be a true reflection of the ability of the cohort though. Rather, outcomes were considered to be artificially increased due to the manner in which students were taught and prepared for assessments (see also Chapter 2). While such approaches are likely to exist in a system of single examination entry, their effects may be heightened when students sit multiple assessments throughout the academic year, since there are more opportunities to train and coach students for the assessments. The greater availability of assessments also provides additional opportunities for schools to receive feedback on their students' performance, which can be used to target future teaching and examination entries (see Chapter 5).

The manner in which students are prepared for assessments is one factor that is likely to influence students' performance on a given assessment in the presence of EMEE. The effects of this are likely to differ though depending on individual circumstances (see Chapter 5). For example, students who are exposed to additional teaching between entry attempts might improve their performance through EMEE, while students who lose motivation might perform poorer in a later entry attempt. Such differences are likely to represent legitimate changes in performance (i.e. it would be expected that additional teaching might contribute to improved performance). However, there are also other factors that can influence students' outcomes due to the nature of assessment, and one of these is the presence of measurement error (see full discussion below). The concept of measurement error was not discussed by respondents in the qualitative phase of this research (though this is not necessarily surprising; see below), yet it is another factor that is likely to impact on the outcomes that students achieve – both in a system of single

examination entry and in the presence of EMEE. The following section outlines the concept of measurement error and discusses the possible implications of measurement error for outcomes in GCSE mathematics in the presence of EMEE. This underpins the analyses in this chapter and Chapter 7, and is considered further in Chapter 8 when the implications of EMEE for maintaining examination standards is discussed.

Measurement error in assessment

As outlined above, there are a number of factors that can influence outcomes on a given assessment. This includes factors that might legitimately affect performance such as preparedness and motivation for the assessment, as well as other factors relating to the nature of assessment. One of the latter factors is referred to as 'measurement error' or 'assessment unreliability'. Measurement error has been defined by Gardner (2013) as '*any variation from the mark, score or grade that characterises the student's 'true' capability in the aspect of their performance that is being assessed*'. It is inherent in assessment (Newton, 2005a) since assessments cannot be perfectly reliable instruments (and should not be interpreted as such; Harlen, 1994; Newton, 2005a; Wiliam, 2001).[79] Essentially though, the presence of measurement error means that if a student sat an assessment several times they would not achieve the same grade each time, even if the same amount of learning had taken place (Haertel, 2006; Wiliam, 2001). Furthermore, any differences in grade would be random. While there are implications of measurement error on any single testing occasion, there are also implications in the presence of EMEE. First we consider the concept of measurement error, before turning to consider the possible implications of measurement error in the presence of EMEE.

[79] Reliability is considered here as the extent to which assessment outcomes would be consistent on multiple occasions (see Haertel, 2006, for further discussion of reliability).

Measurement error

The concept of measurement error (or assessment unreliability) is widely accepted by assessment experts, but less well known in the public domain (see Chamberlain, 2013; Gardner, 2013; He, Opposs and Boyle, 2012; Johnson, Burrough, Ziff and Collins, 2012; Murphy, 2013; Newton, 2013). This is perhaps not surprising (in the UK at least), since there is little discussion or reporting of the uncertainty around assessment results (Bradshaw and Wheater, 2013; see also He and Opposs, 2012). Furthermore, the lack of understanding is likely to be compounded by the terminology that is used – the use of the word 'error' suggests that a mistake has been made, yet the word error here refers to an error in the statistical sense to indicate that there is variability in students' scores (Newton, 2005a). This lack of understanding can be problematic, however, since it can lead to unrealistic expectations about the reliability of assessments and a misinterpretation of their outcomes (Bradshaw and Wheater, 2013; He and Opposs, 2012; Newton, 2005a Wiliam, 2001). Nonetheless, the complexities outlined here suggest that respondents in the qualitative phase of this research were unlikely to have been aware of (or familiar with) the concept of measurement error, and hence would not have considered the implications of this for outcomes in the presence of EMEE.

Measurement error refers to the factors in an assessment that randomly influence outcomes – and this notion of randomness is key. It has therefore been distinguished from other factors that might influence outcomes that are considered to be preventable – for example, human errors that could be avoided (and in this sense are actual errors). Preventable errors might include an error in a question paper, an error in adding up a student's marks across multiple questions, or an error in marking by an errant examiner (Winkley and Cresswell, 2012). When discussing assessment unreliability with stakeholders, Chamberlain (2013) noted the difficulties that some respondents had when trying to disentangle these different sorts of error from measurement error, despite human

errors and measurement error being introduced separately as stand-alone topics. This is perhaps not surprising given the complexities involved.

The presence of measurement error in assessments means that in any single examination attempt a student may randomly achieve a score that does not reflect their 'true' score (the definition of the 'true' score is discussed below). For GCSEs (and A levels), students' scores are converted to grades based on grade boundaries (or cut scores) that are set through the process known as awarding (see Chapter 1). Thus, while measurement error may result in a student achieving a score that does not reflect their 'true' score, there are also possible implications for grades. This has been considered in a body of work exploring classification accuracy, that considers the extent to which a student's observed classification (or grade) aligns with those based on a student's 'true' score (Lee, Hanson, and Brennan, 2002; Livingston and Lewis, 1995) – i.e. it considers the reliability of the test (with reliability defined as the extent to which a given student would get the same result across multiple attempts of the same test; see Haertel, 2006, for further discussion of methods for estimating reliability). A number of recent studies have considered classification accuracy for a variety of English assessments (e.g. Wheadon and Stockford, 2010), suggesting a range of classification accuracy figures (see He and Opposs, 2012, for an overview). These studies have typically considered the percentage of students that are thought to be correctly classified on one testing occasion. Thus, although there are parallels between this work and the issues relating to EMEE, the interest in this thesis lies in situations where students enter on multiple occasions. The literature relating to classification accuracy is therefore not discussed further here.

True scores

Before considering the implications of measurement error for examination outcomes in the presence of EMEE, it is necessary to consider the concept of a 'true' score. In this thesis, 'true' score is operationalised in the context of classical test theory (CTT; or true score theory, see Johnson and Johnson, 2012), that dates back to early work by Spearman

(1904a; 1904b). Accordingly, a student's 'true' score on an assessment is considered to comprise their observed score and an error term (Lord and Novick, 1968), with only the observed score being known. As outlined by Haertel (2006), the true score is considered to reflect the characteristics of a test taker that remain stable across multiple testing occasions and can therefore be considered as '*systematic influences on test performance*', while error is considered to reflect random characteristics that are unstable over repeated testing occasions. Thus, the true score is theoretically stable across tests (but see discussion below), while error can vary, thereby causing differences in the observed score (Brennan, 2001). A student's true score would therefore be given by averaging their scores across an infinite number of tests, meaning that the error is ultimately cancelled out. Although there are alternative operationalisations of the 'true' score (see Baird, Béguin, Black, Pollitt and Stanley, 2012), the approach outlined by CTT is adopted in this thesis since it provides an interpretive framework within which we can consider the implications of measurement error in the presence of EMEE.

This leads to one further point to consider – that is, the extent to which 'true' scores can be considered stable over time. Much of the literature relating to 'true' scores and assessment reliability in the UK has focused on whether students are likely to be correctly classified on a single testing occasion (see He and Opposs, 2012, for an overview). As such, it can be assumed that students have one 'true' score that is relevant to that testing occasion. In EMEE, however, we are interested in students entering on multiple occasions, possibly to different assessments at different times. This raises a question of whether a student's 'true' score remains static over time and across different testing occasions (and on different assessments), since the implications of EMEE are likely to differ.

According to the assumptions of classical test theory (as outlined by Haertel, 2006), an individual's true score is considered stable across multiple *test forms*. However, this assumption is not necessarily applicable in EMEE: in situations where students enter

multiple assessments in the same examination series it might be reasonable to assume that their 'true' score is the same (or at least highly correlated), since they are essentially entering multiple forms of a test at the same time. However, it seems unlikely that this assumption would hold if students are entering at different times, though this might depend to some extent on when the entries are made relative to one another (see Clauser and Wainer, 2016). If only a small amount of additional learning has taken place the 'true' score may remain static, but over a longer time period this is unlikely to be the case. Furthermore, although the 'true' score might change to some extent, the grade associated with this 'true' score might remain the same, since each grade relates to a range of marks (or scores) on the assessments.

In his chapter on reliability, Haertel (2006) discusses the ambiguity in how the 'true' score is defined within classical test theory, and the reality that an individual's performance on a test can and does change over time. Haertel (2006) notes that for some purposes the true score would be considered to be the same over multiple tests, while for others the true score might change over a longer or indeed shorter period of time. Thus, in this thesis it is not assumed that a student's 'true' score is necessarily static or different on every testing occasion in the context of EMEE, since both instances may occur. This means that students might legitimately achieve a different grade on different testing occasions through genuine improvements in performance, but equally they might not. However, it is considered that EMEE has implications for examination outcomes either way, and that is the key point here.

Measurement error and multiple entry

As outlined above, assessment unreliability (or measurement error) is random (Wiliam, 2001). This means that where a cohort of students sits an assessment on a single occasion, some students are likely to perform above their 'true' score and others below their 'true' score. It is therefore reasonable to assume that on a single testing occasion unreliability will average out: that is, the number of students achieving below their 'true'

score is likely to be comparable to the number of students achieving above their 'true' score (Wiliam, 2001). However, the implications in the presence of EMEE are different, and this is because students retain their best grade as their final grade. Let us unpack this further.

If we were to ignore the possible effect of factors such as test preparedness and motivation on student performance and assume that each student's 'true' score is the same across multiple entries, we could surmise that a student entering on multiple occasions would likely achieve different grades on at least some of these entry attempts – sometimes they may under-perform relative to their 'true' score, and sometimes they may over-perform – and this would likely be converted to different grades. In the context of EMEE, the cancelling out effect described previously cannot occur across the whole cohort of students though, and this is because students retain their highest grade as their final grade, which in the presence of EMEE *might* be a higher grade than their 'true' score would suggest (students would be less likely to retain a grade that was below what their 'true' score might suggest since on repeated testing occasions it is unlikely that they would always achieve a lower score due to the random element of measurement error). Across an entire cohort of students outcomes *might* therefore rise, potentially above what might be expected if each student performed close to their 'true' score or if under- and over-performing students cancelled one another out.

This effect was illustrated by Wheadon (2010) who simulated the impact of re-sitting on the fairness of grading. Wheadon (2010) showed that re-sitting increased the percentage of 'false positives' – that is, students achieving a higher grade than their 'true' score would suggest (Hanson and Brennan, 1990; see also Cizek and Bunch, 2007). Furthermore, as the number of re-sits increased, the chance of students achieving a higher grade than their 'true' score would suggest also increased. A similar effect was shown by Millman (1989) in relation to mastery assessment. Millman (1989) plotted the probability of test takers of different levels of mastery passing an assessment requiring 70% mastery

according to the number of entry attempts made. Again, as the number of entry attempts increased the probability of students passing also increased. This effect was particularly acute for those just below the required level of mastery (i.e. those considered to have 65% mastery): on a single testing occasion only around 15% of these students would be expected to pass, but this increased to around 50% after five entry attempts and 80% after 10 attempts. Furthermore, while the percentage pass rate for students with lower levels of mastery (60%) always remained low, this increased from less than 5% to around 15% after 10 attempts. This demonstrates how, for students below the required level of mastery, the practice of entering on multiple occasions increases their change of passing and achieving a 'false positive'. The possible implications of this for different types of assessments are considered by Cizek and Bunch (2007), so are not discussed here. Instead, we consider the parallels with EMEE.

If we were to assume that students enter GCSE mathematics on multiple occasions (but have the same levels of test preparedness and motivation each time), then these students would be more likely to receive a final grade that is higher than their 'true' score would suggest, due to the error that is present in assessment. This could have implications for overall outcomes for a given cohort, that would be likely to rise. It is also worth noting at this point that repeated entry was shown by Wheadon (2010) to reduce the number of 'false negatives' – the number of students achieving a lower grade than their 'true' score would suggest (see also Cizek and Bunch, 2007). This might be considered a positive outcome, since students who randomly perform below their 'true' score on their first attempt would be less likely to be penalised in terms of their final grade. However, while an outright ban on re-sitting is not being advocated here because of this, the opportunity for EMEE to increase outcomes must be considered alongside the possible effects on outcomes and how grades can be interpreted (see also Clauser and Wainer, 2016). This is considered further in Chapter 8.

If it is therefore assumed that the practice of EMEE results in a greater number of 'false positives' – students achieving a grade above what their 'true' score would suggest – then there are several potential implications of this, at both the student and cohort level. The potential for students to achieve a higher grade through EMEE raises questions around how grades can, or indeed should, be interpreted, given the context in which they were achieved. Grades are used to provide an indication of what a student knows and can do and can have implications for accessing further education and employment (see Chapter 1). The presence of measurement error and EMEE might therefore raise questions around the extent to which one can generalise from a student's grade to their ability, and whether the way in which students achieved a grade should be considered when interpreting what that grade means (see Chapter 8 for further discussion).

This issue was touched upon in Chapter 5 (although in the context of students being trained to sit assessments, not in the context of measurement error influencing outcomes): respondents thought that students who had been coached to achieve a particular grade through EMEE would be less proficient than those who had not followed this route, even if the exact same score had been achieved. In these cases, respondents thought that there were possible implications for allowing students access to further education in mathematics and the student's ability to cope with the demands of further study (see Chapter 5) – real concerns for both students and schools.

There are also possible implications at the cohort level, and this is the focus of the current and following chapter. If individual students achieve a higher grade than their 'true' score would suggest through EMEE, then this is likely to result in overall outcomes increasing for a given cohort, since there is no opportunity for students that under- and over-achieve (relative to their 'true' score) to average out. Thus, there are possible implications for the way in which outcomes for a given cohort of students should be interpreted, particularly when making comparisons between different years. For example, while a change in outcomes for a given cohort (e.g. 16-year-olds) from one year to the next might be

considered to reflect a change in performance (as is often the case when GCSE results are reported; see Chapter 1), this might actually reflect a higher proportion of students entering on multiple occasions. This creates challenges when interpreting national outcomes, or when making comparisons between different cohorts of students, since there are likely to be differences due to the manner in which students entered the assessments, rather than due to the characteristics of the students (see Chapter 8 for further discussion).

The presence of EMEE therefore also raises questions concerning the way in which standards are (or indeed should) be maintained, an issue that is considered in full in Chapter 8. First though, the current and following chapter consider the possible implications of EMEE for examination outcomes. In these chapters the focus is on the overall outcomes at the cohort level, rather than the implications for individual students (though the two issues do, of course, interact).

Aims and rationale

The over-arching aim of the current and following chapter is to explore the extent to which EMEE might impact upon outcomes in GCSE mathematics. The chapters also reflect upon the extent to which this has implications for interpreting the grades that students achieve. The analyses are two-fold. First, analyses of empirical data are undertaken using data from the NPD (see below) – this forms Chapter 6. Second, simulations explore the extent to which EMEE can influence outcomes due to the presence of measurement error in assessment – this forms Chapter 7.

The two chapters complement one another. This chapter uses empirical data to consider the use and implications of EMEE in practice. Initially, this includes a consideration of the prevalence of multiple entry and the characteristics of students following different entry routes, before the potential impact on outcomes is considered. The former analyses provide the necessary context for the latter analyses that focus on the implications for

outcomes, as without this we would be trying to consider the implications of multiple entry without fully understanding the behaviour itself. The analyses therefore address RQ3: what characteristics at the school and student level are associated with multiple entry to GCSE mathematics assessments? and RQ4: what are the implications of EMEE for examination outcomes in GCSE mathematics assessments?[80] In doing so, a number of subsidiary research questions are considered, focusing on data from the 2012 – 2013 academic year (see discussion below):

RQ3a) What was the balance of single and multiple entrants within the overall cohort of students?

RQ3b) On how many occasions did multiple entrants certificate?

RQ3c) What are the characteristics of students following alternative entry routes through GCSE mathematics?

RQ4a) How does EMEE influence examination outcomes when considering a student's best grade vs. all grades or their average grade?

RQ4b) How does multiple entry in the same examination series influence outcomes?

Limitations of the analyses

Before considering the methodology a number of limitations are discussed that should be borne in mind when interpreting the findings. In particular, there are limitations to the conclusions that can be drawn, particularly with regards to the final two research questions. As indicated in Chapter 5, schools have highly individualistic approaches for entering their students to GCSE mathematics, meaning that it cannot be assumed that students are fully (or equally) prepared for each examination attempt. For example, some schools might enter students early despite them not being fully prepared for the

[80] Note that RQ4 is considered further in Chapter 7.

assessment with the intention to re-enter them in a later series after further tuition, while others might enter students early with greater preparation then re-enter them in a later series with little teaching time in-between, particularly if the students have already achieved their target grade (see Chapters 2 and 5). These differing levels of preparedness are likely to influence the grades that students achieve on different entry attempts. In the former case, students may perform poorly on an early attempt then improve at a later date, while in the latter the converse may be true. Indeed, some schools may enter students to examinations early knowing that they will perform poorly and that the grade achieved will not be their final grade. However, this approach provides examination practice and allows schools to target the areas where their students performed poorly through further tuition (see Chapter 5).

The use of different entry and teaching strategies means that any differences identified in outcomes when comparing students' grades on different entry attempts are likely to be due to a number of factors. This includes the presence of measurement error as discussed above, but also other factors such as improvement over time, maturation effects, differing levels of preparedness and motivation, and the extent to which students are 'trained' for the assessment. Some of these factors, in particular the extent to which students are 'trained' for assessments, were considered by respondents in the qualitative study to artificially inflate national outcomes and not provide a true reflection of a student's ability. Other factors, however, such as natural improvement over time (possibly due to maturation effects) are likely to be entirely legitimate reasons for performance to change. Consequently, there are likely to be a combination of factors that influence outcomes – some that may legitimately reflect changes in student performance (e.g. further tuition) and some that might not (e.g. measurement error) – and it is not possible to disentangle these given the data available.

While disentangling the effects of measurement error from the other factors identified above is not possible in many of these analyses, the final set of analyses in this chapter

provides some insight, by exploring differences in performance for students entering two specifications in the same examination series. To a large extent, this comparison should minimise the effects of maturation, preparedness and motivation. However, such effects are unlikely to be completely eradicated, even when comparing performance at the same time point since students are entering different specifications and may have been better prepared for one or the other. Given this, the possible impact of measurement error is explored further in Chapter 7, where the impact of multiple entry on outcomes is modelled using simulations. This complements the findings reported in the current chapter, by offering an additional perspective on the possible effects of EMEE on outcomes. Thus, the empirical analyses in this chapter also provide the context for the simulations in Chapter 7.

Methodology

To explore the research questions outlined above, access to the NPD was arranged via the DfE. This database contains detailed information for all students based in schools and colleges in England. This includes examination results and attainment at different key stages of education, alongside a number of demographic variables. Importantly for these analyses, it provides a record of each student's GCSE examination attempts in every subject with the corresponding grade and examination series (at the overall qualification level), for those students completing KS4 (i.e. completing Year 11) in a given academic year.

Data was accessed for the 2012 – 2013 academic year, since this was the last year prior to the announcement that only a student's first entry attempt would count towards school performance measures (see Chapter 1). This change to accountability measures had a significant impact on entry patterns, reducing the prevalence of EMEE considerably the following year (see Chapters 1 and 8). Consequently, EMEE was likely to have been at a peak during the 2012 – 2013 academic year. Data was only accessed for one academic

year. While making comparisons between years would have provided some additional insight, particularly during times of policy change, such analyses would have been limited since the availability of assessments has changed considerably each year, meaning that any trends would be difficult to interpret.

At the time the data was collected there were a number of GCSE mathematics specifications available across multiple awarding bodies (see Chapter 1). This included linear specifications, modular specifications, pilot specifications and international GCSEs (also known as level 1/2 certificates; see Chapter 2). For the purpose of these analyses, entries to the main linear and modular GCSE specifications and international GCSEs[81] from all UK awarding bodies were included (see Chapter 1): these qualifications each counted towards performance tables (at the time the analyses were conducted), hence their inclusion in the analyses. Entries to the linked pair pilot mathematics specifications were excluded.[82] The linked pair specification required students to sit two mathematics qualifications – applications and methods – with a student's best grade counting towards performance measures. However, this grade only counted if students had certificated in both qualifications, meaning that including this qualification would have complicated the analyses significantly. Furthermore, since this was a pilot specification the entries were small (each awarding body was restricted to a maximum of 100 centres while the qualification was in the pilot phase), suggesting that excluding entries for this qualification would not impact on the results.

Entry attempts where students did not receive a grade (i.e. they were absent from the examination) were also excluded.[83] There are likely to be a number of reasons why

---

[81] The AQA certificate in further mathematics was excluded since although this is a level 2 qualification, the content and standard is different to other GCSE mathematics qualifications.

[82] Note that if students entered the main GCSE mathematics qualifications and the linked pair pilot qualifications their entries for the main GCSE would be included in the analyses.

[83] Note that candidates who did not achieve a grade are coded as 'ungraded' in the NPD, hence are distinct from candidates achieving a grade U.

students are absent from an assessment (see Chamberlain, 2008). This includes illness (where the student intended to enter but did not), and administrative errors (where the student was entered but never intended to sit the assessment).[84] Although excluding absentees from these analyses might under-estimate the number of times that some students intended to enter GCSE mathematics (since the majority of students were likely to be absent for the former reason), including them would assume that all students who were absent intended to enter. This might over-estimate the prevalence of multiple entry so the more conservative approach was adopted. This also has the benefit of reflecting the situation in practice (i.e. students were only included if they actually sat the assessment), rather than speculating upon what might have been.

Students were removed from the data if they had duplicate entries in the NPD, since it was not possible to determine which data were correct. However, since duplicates represented a minor percentage of the total number of students (0.06%) their exclusion is unlikely to influence the findings. Students were also excluded if they were not aged 15 at the start of the 2012 – 2013 academic year. Students typically turn age 16 during Year 11, meaning that the excluded students were likely to be part of other year groups but were sitting GCSE mathematics in the 2012 – 2013 academic year (e.g. they were certificating at an earlier or later age than expected; see Chapter 4 for analyses of early entry). The final data comprised 595,240 unique students who completed Year 11 during the 2012 – 2013 academic year and had sat GCSE mathematics at least once. For each student a record of every examination attempt in GCSE mathematics was available, alongside other demographic variables.

It is worth noting here that, although the NPD provides a record of each student's examination entry attempts (including the examination year and series), there is no

[84] This is likely to be the case when students are entered to a subsequent examination series before receiving their grade from a previous series, then not withdrawn by their centre.

distinction between the three winter examination series (November, January and March) that were available in 2012 – 2013, and these were all recorded as 'winter' in the data. Thus, it was not possible to order students' entry attempts chronologically or be certain when students sat assessments if they entered multiple times in a winter series. For example, a student certificating in the winter 2012 series may have entered in November 2012, January 2013 or March 2013. In the data, 61,538 students (10.34% of the overall cohort) first certificated in a winter series and re-certificated the same winter, so it was not possible to be certain which attempt was their first. Thus, these students were excluded from any analyses that considered a student's first grade. Similarly, students were excluded from such analyses if their first attempt involved multiple entries in the same summer series since again, it was not possible to determine which their first attempt was. These students represented a small percentage of the overall cohort though (8,401 students; 1.41%).

Results

Single or multiple entrants?

To address the first two subsidiary research questions, Table 14 shows the number of entries to GCSE mathematics per student prior to their completing Year 11 in the 2012 – 2013 academic year. As discussed previously, students are typically expected to certificate in each GCSE subject once at the end of Year 11 (see Chapter 1), so it is striking that only half of the students in this cohort followed this route (see also DfE, 2013b). This suggests that single entry, the route traditionally taken, was not the 'norm' during this period. Rather, large numbers of students entered on multiple occasions.[85] Of these multiple entrants, the majority were also likely to be early entrants (to allow an

[85] Note that these figures only include the number of entries that students made up to them completing Year 11, so any entries in post-16 education are not included.

opportunity to re-enter), unless students certificated multiple times in their final examination series (i.e. the summer 2013 series).

Table 14. *Number of students entering on single or multiple occasions*

Entry attempts	Number of students	Percentage of students
1	296,337	49.78
2	186,740	31.37
3	78,024	13.11
4	24,026	4.04
5	6,821	1.15
6	2,084	0.35
7	649	0.11
8	357	0.06
9	145	0.02
10	46	0.01
11	8	0.00
12	3	0.00
Total	595,240	100.00

While multiple entry was clearly a popular route during this period, the number of occasions on which students entered differed fairly significantly, ranging from those who entered twice to those who entered on up to 12 occasions (see Table 14). Within these groups, students were also likely to have followed different entry routes, since multiple entry could refer to students entering in the same or different examination series, with the same or different awarding body, and to the same or different specification. This is illustrated in Table 15 that shows a cross-tabulation of the examination series in which students entered if they entered twice. As shown, the number of entry routes is vast (40 routes), although some are much more popular than others. For example, entering part-way through Year 11 and again at the end of Year 11 is the most popular route (121,395 students), while entering twice in summer 2012 is significantly less popular (44 students).

Table 15 illustrates the large number of entry routes for students entering GCSE mathematics twice before the end of Year 11, yet this does not account for the awarding

body or specification that students entered – or whether students entering in a winter series entered in the same or different series (i.e. November, January or March) – which would increase the number of routes significantly. This highlights the complexity and vast array of entry routes that students took through GCSE mathematics at this time (see also Chapter 5). Furthermore, it suggests that focusing upon high level trends throughout this chapter rather than attempting to categorise and compare students by their specific entry route is most pragmatic, particularly when it is considered that as the number of entry attempts increases, the number of permutations increases rapidly.

Table 15*. Cross-tabulation of examination series for double entry students*

		Second examination series								
		2009W	2010S	2010W	2011S	2011W	2012S	2012W	2013S	Total
First examination series	2005S	0	0	0	0	0	1	0	0	1
	2006S	0	0	0	0	0	1	0	0	1
	2007S	0	0	0	0	0	0	0	5	5
	2008S	0	0	0	1	0	1	3	16	21
	2009S	2	2	0	0	0	0	1	4	9
	2009W	0	0	0	0	0	0	2	2	4
	2010S	0	0	2	23	2	22	10	24	83
	2010W	0	0	1	14	5	3	31	54	108
	2011S	0	0	0	2	282	213	507	610	1,614
	2011W	0	0	0	0	250	1,358	2,891	2,862	7,361
	2012S	0	0	0	0	0	44	11,851	14,812	26,707
	2012W	0	0	0	0	0	0	21,624	121,395	143,019
	2013S	0	0	0	0	0	0	0	7,806	7,806
	Total	2	2	3	40	539	1,643	36,920	147,590	186,739

Note: S refers to the summer examination series (June) and W to the winter examination series (November, January or March). One student was excluded from this table as their first entry was in summer 2001 which seems unlikely to be accurate.

In terms of the number of entries, multiple entrants typically entered on two or three occasions, collectively accounting for 44.48% of all students completing Year 11 in summer 2013 (see Table 14). However, the number of students entering on a greater number of occasions was not insignificant. Indeed, the number of students entering on

four or even five occasions was large enough to suggest that these were not isolated incidents or unusual cases. This data reflects the qualitative findings in Chapter 5, where a number of respondents reported having entered their students four or five times prior to their completing Year 11. While this does not necessarily suggest that the schools included in the qualitative study were representative of all schools, it does suggest that the approaches identified in the qualitative strand were unlikely to be unique to those schools.

At the extreme, three students entered GCSE mathematics a total of 12 times, with each of these students entering repeatedly throughout Years 9, 10 and 11, sometimes to multiple specifications in the same examination series. While this, and the proportion of students entering on more than five occasions, represents a small proportion of the overall cohort, it is not insignificant, particularly given the potential consequences that this level of multiple entry could have for students, schools, and the wider assessment system. Any form of multiple entry could impact on students by affecting their motivation and engagement which, in turn, could impact upon the outcomes achieved and future progression (see Chapter 5). As the number of entries increases, these effects are likely to be exacerbated.

Multiple entry is also likely to have implications for schools. Repeatedly entering students to assessments can cause disruption and reduce time for classroom teaching, since each examination series will likely be accompanied by intense periods of revision, an issue identified in the qualitative phase of this research (see Chapter 5). This might impact on the subject in question (in this case, mathematics), as well as other subjects that might be deemed ‘less important’ – i.e. those where the outcomes have less significance for performance measures.

Alongside these educational effects, another implication at the school level is the cost. The prevalence of multiple entry shown here will have significant cost implications for individual schools and across the whole examination system, an issue highlighted elsewhere (see DfE, 2013b). Considering that each entry to GCSE mathematics cost

around £30 in 2012 – 2013, it can be calculated that for the 2012 – 2013 cohort the dual entrants alone increased the money spent on examination fees for GCSE mathematics significantly, by around £5.6 million. When considering all entry attempts, this figure rises to over £13 million. While some may argue that this is good news for examination boards in terms of boosting their revenue,[86] it is less so for the schools who are likely to be paying. However, it does raise an interesting question around when multiple entry might be considered 'good value for money' by schools, if it boosts outcomes and avoids the negative implications of being below floor standards, for example (see Chapter 1).[87]

What are the characteristics of students following alternative entry routes?

The number of entry routes through GCSE mathematics (identified in Table 14) and the manner in which schools entered students to assessments (see Chapter 5) suggests that there are likely to be differences in the type of students following each route. Indeed, evidence that early entry was associated with particular types of students has already been presented in Chapter 4 of this thesis (see also Chapter 2). Similar differences might also exist both between single and multiple entrants, and between those entering multiple times but on different numbers of occasions. This assertion is supported by the findings in Chapter 5, where multiple entry was often reported as being concentrated amongst the grade C/D borderline students.

The following section explores this further and addresses the third subsidiary research question: RQ3c) what are the characteristics of students following alternative entry routes through GCSE mathematics. Several analyses are presented. First, entry route is considered in light of the initial grade that students achieved. If, as discussed in Chapter 5, EMEE is concentrated amongst the grade C/D borderline students, it is reasonable to

---

[86] Note that the awarding bodies were vocal in their concerns about EMEE (e.g. Hall, 2012) and there is no suggestion that they supported EMEE to boost their revenue.

[87] Note that this point ignores any negative effects of EMEE and is focused on the final grade achieved.

hypothesise that multiple entry will be most prominent amongst students who achieved lower than a grade C initially, since there will be greater impetus to improve this grade in a later attempt. Second, the characteristics of students following different entry routes are explored, in terms of their performance outcomes and prior attainment. Third, these relationships are modelled using logistic regression, where the association between a student's entry route (single or multiple entry) and a number of student and school level characteristics are explored.

Multiple entry for students with lower prior attainment

The prevalence of single or multiple entry amongst students according to the first grade that they achieved is presented in Table 16 (note that for single entrants this is also their final grade). To ensure that a student's 'first' grade is accurately captured, students were excluded if they had multiple entries in the first series that they certificated – be it the summer or winter – since it would not be possible to determine which of these was the first or second grade. The excluded students represented a relatively small percentage of the overall cohort (11.75%).

As shown in Table 16, students achieving below a grade C on their first attempt were more likely to re-enter in a later examination series than those achieving a grade C or above; indeed, the prevalence of multiple entry is greatest for those students achieving a grade D initially (see also Chapter 2). This is not particularly surprising, since these students are arguably most likely to achieve a grade C on a subsequent entry attempt, and this is the threshold for school performance measures (see Chapter 1). There is also a trend for students who achieved below a grade C initially to make a greater number of entry attempts overall, on average (see Table 16). Again, this is not surprising, since these students are likely to be entered repeatedly to provide multiple attempts at achieving a grade C (see Chapter 5).

Despite these trends, there are also instances of higher achieving students entering on multiple occasions. For example, around half of the students achieving a grade C initially

re-entered in a subsequent series. This may stem from students trying to improve their grade – perhaps due to moving between tiers, attempting to achieve a target grade, or to satisfy a progress measure (see Chapter 5) – or simply because the whole cohort of students were re-entered in their school, regardless of a student's initial grade. Indeed, there are some instances of students who initially achieved a grade A* re-sitting in a later series (see Table 16).

The majority of students that achieved a grade A* in their initial attempt and re-entered (79.08%) sat a GCSE and an international GCSE, suggesting that these different qualifications could have been used as enrichment opportunities (the GCSE qualification typically came first). However, the remaining students re-sat another GCSE specification, despite achieving an A* on their first attempt. In some instances, students sat the same modular specification on both occasions so it is possible that they were re-entered by their centre but did not sit the final module in their second attempt. This appears likely, since students following this route typically performed much poorer on their second attempt (although it is not possible to be certain since partially absent students are not flagged in the NPD).[88] Where this is not the case though, re-entry may have been used to ensure that students continued to study mathematics throughout Year 11. There are questions around the appropriateness of this though, since these students may have benefited more from acceleration or enrichment opportunities (see Chapter 2). In some instances, this may have been available and students may have studied advanced material alongside re-entering GCSE mathematics. However, the appropriateness of re-entering at GCSE still remains questionable, given their early success.

[88] Partial absentees are candidates who have certificated but have not completed all the required components of a specification. They are typically flagged in awarding body data and are excluded from the predictions generated for awarding purposes.

Table 16. *The prevalence of multiple entry, mean entry attempts and maximum entry attempts by first grade*

First grade	% Multiple entry	Multiple	Single	N	Mean	Max
A*	0.87	392	44,617	45,009	1.01	3
A	13.85	9,115	56,702	65,817	1.16	7
B	26.42	23,877	66,482	90,359	1.33	7
C	46.66	68,318	78,090	146,408	1.60	10
D	77.22	50,569	14,921	65,490	2.23	10
E	73.88	31,226	11,039	42,265	2.29	11
F	67.24	22,939	11,176	34,115	2.11	12
G	63.65	14,869	8,491	23,360	1.98	10
U	61.38	7,659	4,819	12,478	2.01	9

Multiple entry and lower outcomes

As shown above, multiple entry appears to be concentrated amongst the C/D borderline students and the weaker students. To consider further the characteristics of students following different entry routes, outcomes in GCSE mathematics and prior attainment data were considered according to the number of times that each student certificated. Considering outcomes first, Table 17 shows the cumulative percentage outcomes at the key GCSE grades (see Chapter 1) according to the number of entry attempts each student made.[89]

When taking into account the best (or final) grade that each student achieved across all entry attempts, it is evident that the outcomes differ fairly significantly according to the number of entry attempts made. This is particularly the case when comparing single entrants with any of the multiple entrants, with the former tending to out-perform the latter across the grade range. In particular, the single entrants performed well at the highest grades of A* and A. Although multiple entrants tended to perform similarly regardless of the number of entry attempts, there is some evidence that as the number of entry

[89] Students who entered on five or more occasions were combined since the numbers were relatively small, particularly as the number of entries increased.

attempts increased, outcomes decreased, particularly at the higher grades. This is not to say that multiple entry itself has necessarily resulted in lower outcomes, since there are likely to be a range of factors that have influenced these findings. This includes the ability of the students following each entry route and the way in which students are prepared for assessments and the effectiveness of this.

Table 17*. Cumulative percentage outcomes (best grade) by number of entry attempts*

Entry attempts	A*	A	C	F	Number of students
1	15.06	34.19	82.98	95.51	296,337
2	2.31	10.37	66.79	94.09	186,740
3	1.13	7.27	60.25	95.27	78,024
4	0.83	5.80	60.61	96.87	24,026
5+	0.65	4.89	61.21	97.07	10,113
Total	8.41	21.55	73.65	95.11	595,240

To consider the former, prior attainment was used as a proxy for ability and comparisons made between the prior attainment profiles of students certificating on differing numbers of occasions. Prior attainment was defined as a student's average KS2 test level across English and mathematics. This measure has been used by the awarding bodies more recently when generating statistical predictions for setting grade boundaries (see Chapter 1), and has been identified as a reliable predictor of performance at KS4 (Eason, 2010; 2015; see also Strand, 2006). The analysis only included those students for whom KS2 scores in English and mathematics were available in the NPD (Science KS2 test score is not available in the NPD),[90] and excludes students who did not sit both KS2 assessments. The average English and mathematics performance was used rather than just performance in mathematics for two reasons: first, this measure is used as a predictor for

[90] Prior to summer 2015 prior attainment predictions have used the average level across English, mathematics and science. However, evidence suggests that using English and mathematics yields similar outcomes so this is considered a valid alternative (see Eason, 2015).

awarding purposes (see Chapter 1), hence this analysis remains consistent with these procedures, and second, the KS2 levels are coarse measures (see Chapter 4), meaning that only including mathematics would result in limited discrimination between students.

Figure 14 shows students' average KS2 score according to the number of entry attempts made. As illustrated, there is a gradual decline in the percentage of students with the highest mean KS2 scores (KS2 mean = 4.5 and 5.0) as the number of entry attempts increase. In particular, the percentage of students with the highest prior attainment drops off significantly when comparing single entrants with any of the multiple entrants, regardless of how many entries the latter made. These declines are (inevitably) offset by an increase in the percentage of students with mid-range and lower mean KS2 scores.

The prior attainment profile of each group suggests a difference in the type of students that entered once and those that entered on multiple occasions, with the former tending to be the higher ability students. This is likely to provide some explanation for the differences in achievement observed in Table 17. It also concurs with the qualitative findings reported in Chapter 5, where many respondents reported that students around the grade C/D borderline were most likely to enter early and subsequently on multiple occasions, and previous findings (see Chapter 2).

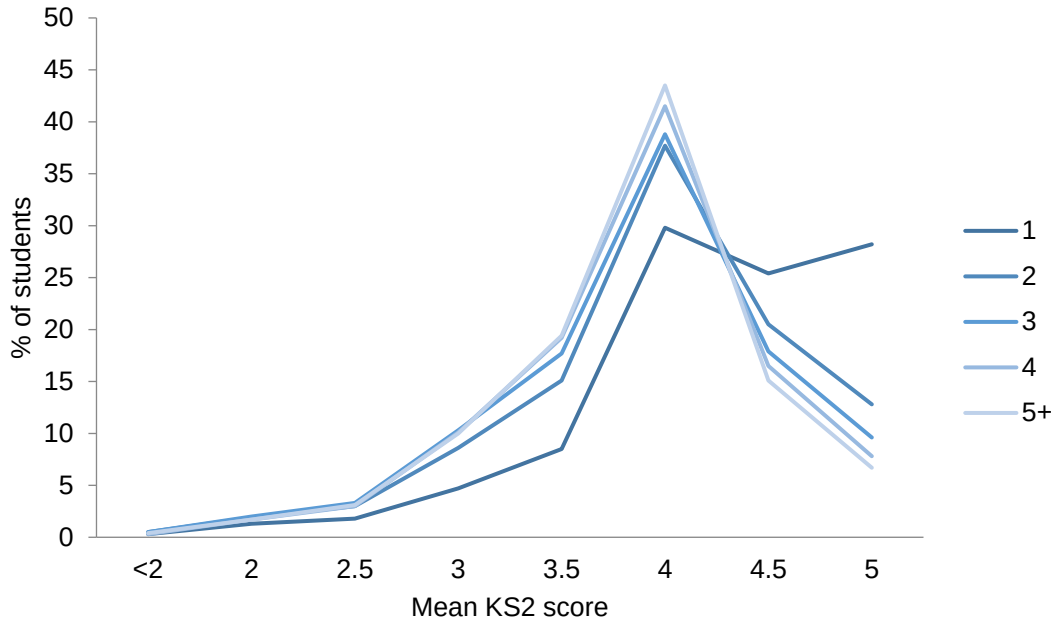


Figure 14. *Mean KS2 profile of students by number of entry attempts*

Predicting single or multiple entry

To explore these relationships further, two logistic regression analyses were conducted.[91] The first regressed entry route (single or multiple entry) on student level variables, and the second added school level variables into the model. At the student level the variables included prior attainment and eligibility for FSM, while at the school level the variables included performance outcomes of the school and the number of disadvantaged students (% eligible for FSM). Prior attainment has been identified as a reliable predictor of performance at GCSE and is used by the awarding bodies to support the maintenance of standards (see Chapter 1; Eason, 2010). Eligibility for FSM and school outcomes have both been associated with early entry, with early entry typically being more prevalent in

[91] Note that for these analyses single level models were used. While it is acknowledged that the data is clustered within schools, multi-level techniques were not used since pursuing school level effects was not the main focus of this thesis. If it were, then further analyses using multi-level techniques could be conducted to check the robustness of the findings and to consider these relationships further.

schools with lower results and a higher percentage of students eligible for FSM (see Chapters 2 and 4).

The outcome variable and eligibility for FSM were coded as binary variables – single or multiple entry, and eligible or not eligible for FSM. Prior attainment was measured as a student's mean KS2 level across English and mathematics, ranging from 0 to 5 (see Chapter 4). School level data were retrieved from the performance table data that is publically available from the DfE. School performance outcomes were measured as the percentage of students within the school achieving five GCSEs at grades A* – C in the 2011 – 2012 academic year (including English and mathematics), and the percentage of disadvantaged students was measured as the percentage of students in each school eligible for FSM (see discussion in Chapter 4 relating to FSM as a measure of disadvantage). The school level data related to the previous academic year to the student level data, since a school's outcomes in the previous year are most likely to influence future entry approaches, particularly if schools are below or close to floor targets. Students were excluded if they were missing any of the above data (17.1% of the overall cohort of students were missing school level data).[92]

The findings from the logistic regression models are presented in Table 18. For both models, the Hosmer-Lemeshow goodness of fit was significant, suggesting poor model fit ($\chi^2 = 8765.20$, df = 8, $p < .001$ and $\chi^2 = 4868.83$, df = 4, $p < .001$, respectively). However, with a large sample size (524,877 and 493,410, respectively) this is not unexpected, since small deviations between observed and expected values will likely yield a significant result. The first model including student level variables was able to correctly classify

[92] Note that this means that any students from schools that are missing data in the DfE performance tables were excluded. An investigation of these schools showed they were typically community special schools, independent schools, independent special schools and non-maintained special schools.

60.6% of cases, while adding in the school level variables increased this figure to 67.0%: 66.6% of single entrants and 67.3% of multiple entrants.

For both models, the analyses revealed a significant association between each of the predictor variables and whether a student entered once or on multiple occasions ($p < .001$). According to the first model, students were more likely to enter on multiple occasions if they had lower prior attainment and were eligible for FSM, assuming that all other variables were held constant. However, given the large sample size, the predictors are interpreted in terms of the odds ratio rather than the significance values. The model suggests that students eligible for FSM were 73% more likely than students not eligible for FSM to multiple enter, after controlling for prior attainment (odds ratio = 1.727). Furthermore, for every unit increase in a student's prior attainment, the odds of entering on multiple occasions decreased by 44% (odds ratio = 0.563). The second model, including school level variables, revealed that students were more likely to multiple enter if they were from schools that had lower performance outcomes and a higher percentage of disadvantaged students, assuming all other variables were held constant. Overall, these findings suggest that multiple entry was an approach to provide weaker students with multiple attempts at achieving a particular grade, most likely a grade C. The association between prior attainment and entry route also suggests that weaker students were more likely to be entered early (see also Chapters 2 and 4), since this would typically be necessary to facilitate multiple entry (unless students entered multiple times in their final examination series).

Although the second model classified a high percentage of cases correctly, around a third were incorrectly classified. This is perhaps not surprising, since the evidence presented in Chapter 5 suggested that while EMEE was typically used for grade C threshold students, there were also other reasons for entering students early and on multiple occasions. For example, 'at-risk' students were likely to be entered early (i.e. mid-way through Year 11) since they were unlikely to be in school and available for assessments at the end of Year

11 (see Chapter 5). Thus, these students were likely to be single entrants, but were not likely to be the higher achievers. Furthermore, the previous analyses in this chapter suggested that higher ability students are, on occasion, entered multiple times, despite achieving a high grade on their early attempt.

Table 18. *Logistic regression predicting single or multiple entry*

Model 1

Variables	Estimated coefficient	Standard error	Odds ratio
Constant	2.322	.016	-
Prior attainment	-0.574	.004	.563*
FSM	0.546	.009	1.727*

* $p < .001$, R^2 = 0.04 (Hosmer-Lemeshow), 0.06 (Cox-Snell), 0.07 (Nagelkerke)

Model 2

Variables	Estimated coefficient	Standard error	Odds ratio
Constant	1.275	.024	-
Prior attainment	-0.466	.004	.627*
FSM	.074	.010	1.077*
School performance	-.005	<0.00	.995*
School disadvantage	.042	<0.00	1.043*

* $p < .001$, R^2 = 0.11 (Hosmer-Lemeshow), 0.14 (Cox-Snell), 0.19 (Nagelkerke)

To summarise, this section has explored the characteristics of students following different entry routes through GCSE mathematics. The findings suggest that entry approaches are not arbitrary: there is an overall trend for weaker and more disadvantaged students to enter on multiple occasions. This reflects the findings in Chapter 5, where some schools reported concentrating EMEE amongst the grade C/D borderline students, and previous evidence (see Chapter 2).

Single entrants

The previous section has focused on differences between single and multiple entrants by exploring the characteristics of students following alternative entry routes. Throughout these analyses the single entrants were therefore treated as a homogenous group, yet

there are also likely to be differences between students that entered once in terms of when these students entered (early entry or at the end of Year 11), their performance, and their prior attainment. For example, early entry was cited as a standalone approach for accelerating the highly able students by respondents in Chapter 5, suggesting that some of the single entrants may have followed this route. Alternatively, early entry was used for 'at-risk' students who were unlikely to remain in education until the end of Year 11 (see Chapter 5). To explore this, the following section considers the outcomes and prior attainment of single entrants according to the examination series in which they certificated. This allows comparisons to be made between single entry students who entered early and those who entered at the end of Year 11.

To provide some context to these analyses, Table 19 shows the series in which single entry students certificated. As shown, the majority of single entrants (68.18%) certificated in the summer 2013 series, the point at which students would typically be expected to sit their GCSEs (i.e. at the end of Year 11; see Chapter 1). The remaining students were early entrants; 25.68% entered in the winter 2012 series (i.e. part-way through Year 11), 5.30% entered in summer 2012 (i.e. at the end of Year 10), and 0.85% entered at an earlier stage (i.e. part-way through Year 10 or earlier). The latter group are likely to represent unusual cases, including those students who certificated at a much younger age than expected (see Chapter 1), so are grouped together throughout these analyses.

Table 19. *Examination series entered for single entry students*

Series	Number of students	Percentage of students
Winter 2011 or before (mid Y10 or before)	2,506	0.85
Summer 2012 (end of Y10)	15,692	5.30
Winter 2012 (mid Y11)	76,096	25.68
Summer 2013 (end Y11)	202,043	68.18
Total	296,337	100.00

Single entrants – outcomes and prior attainment

Considering students' outcomes first, Table 20 shows the cumulative percentage outcomes at the key GCSE grades for students according to the examination series in which they certificated. Although the single entrants performed well collectively compared to the multiple entrants (see Table 17), there are differences depending upon which series they entered. At grade C, students entering at the end of Year 10 or part-way through Year 11 performed similarly, out-performing those entering before the end of Year 10 and those entering at the end of Year 11. Indeed, the students entering at the end of Year 10 and mid-way through Year 11 performed well at this threshold grade, with 91.91% achieving a grade C or above in summer 2012 and 93.31% achieving a grade C or above in winter 2012. This compares to 76.14% of students achieving a grade C or above before the end of Year 10, and 78.47% achieving a grade C or above at the end of Year 11. At the top grades though, students entering at the end of Year 10 or earlier tended to perform the highest, with high percentage outcomes at grades A* and A (see Table 20). Interestingly though, while students entering before the end of Year 10 performed well at the top end, the percentage of students achieving a grade C or a grade F was also relatively low (76.14% and 87.99, respectively). This suggests that this group of students were quite disparate, containing both high and low attainers – those that achieved an A* and those that achieved a low grade (see also Chapter 5).

Despite this, the overall outcomes provide some indication that, of the single entrants, early entry had been used for the more able students, particularly when it is considered that over half of the students that entered a year early (at the end of Year 10) achieved a grade A or above. This far exceeds the outcomes that would be expected from a typical cohort of GCSE mathematics students (in 2016 just under 20% of all 16-year-old entries to GCSE mathematics in the UK achieved a grade A or above; see JCQ, 2016). As such, early entry may have been the most appropriate approach for these students, allowing them time to study additional mathematics qualifications in Year 11 or begin the

A level programme of study early (see Chapters 1 and 2). Furthermore, given the high outcomes for each group of early entrants, it is perhaps not surprising that these students only entered once, since the majority had achieved a grade C or above on their first attempt. This is not to say that these students could not have achieved a higher grade had they re-entered or had they only entered once at the end of Year 11 and had additional teaching time though.

Table 20*. Cumulative percentage outcomes by examination series for single entrants*

Entry attempts	A*	A	C	F	Number of students
Winter 2011 or earlier	28.37	46.49	76.14	87.99	2,506
Summer 2012	33.02	55.58	91.91	96.27	15,692
Winter 2012	11.04	27.08	93.31	97.21	76,096
Summer 2013	15.01	35.05	78.47	94.90	202,043

The differences in outcomes shown in Table 20 for students entering in each examination series are likely to reflect differences in the type of students following each route. As such, consideration was also given to the prior attainment profiles of students entering once but in different examination series. This analysis was restricted to those students with KS2 data for both English and mathematics.

As shown in Figure 15, students certificating a year early (in summer 2012) tended to be high ability students, with 43.23% achieving a mean score of level 5 across English and mathematics tests at KS2. This is not surprising, given the high outcomes achieved by these students at GCSE (see Table 20). A similar trend is observed for the students who certificated in winter 2011 or earlier (i.e. 18 months or more years early): these students also performed well at the top end (see Table 20) and a relatively high percentage of these students achieved the highest mean KS2 score of level 5 (34.14%). The students certificating in the winter 2012 series appeared to be the weakest at the top end in terms of their prior attainment, with a high percentage of these students achieving a mean KS2 score of level 4, the level that students were expected to achieve at the end of KS2. This

is also reflected in the outcomes in Table 20, with these students achieving the lowest percentage of A* and A grades. This suggests that if students entered once, those entering a year or more early were generally the higher achievers who performed well (see also Chapter 5), although there are also instances where this trend is not true. For example, over 10% of students entering two or more years early did not achieve a grade F. This could include those students who are considered to be 'at-risk' and are entered early since they are unlikely to remain in education until the end of Year 11 (see Chapter 5).

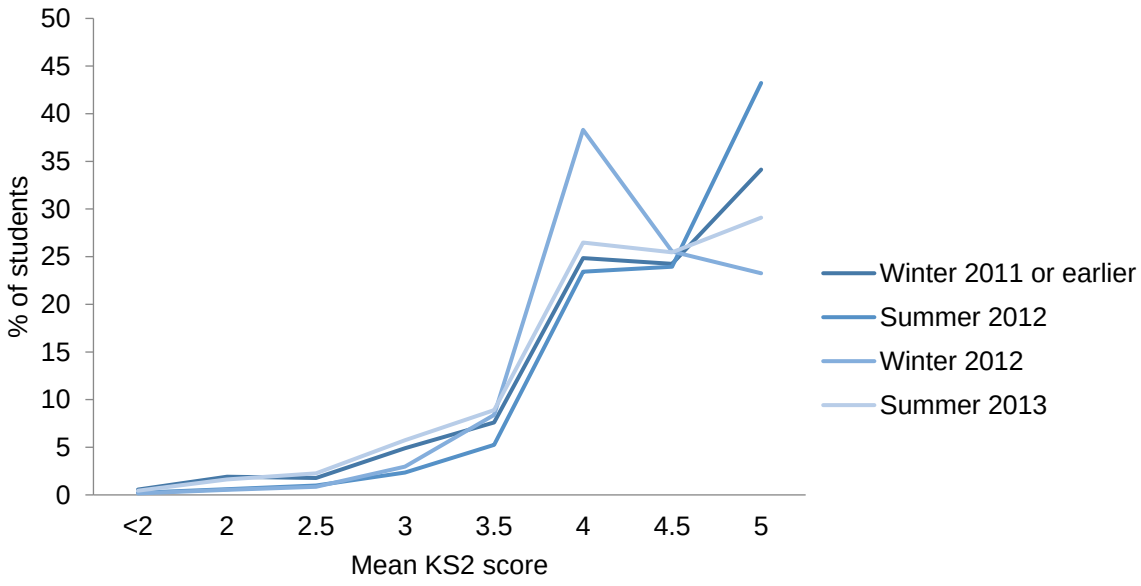


Figure 15*. Mean KS2 profile of single entry students by examination series*

GCSE mathematics outcomes

Thus far, these analyses have explored trends in the type of students following alternative entry routes through GCSE mathematics. The chapter now turns to explore the impact that multiple entry might have on outcomes in GCSE mathematics to address the fourth and fifth subsidiary research questions: RQ4a) how does EMEE influence examination outcomes when considering a student's best grade vs. all grades or their average grade?,

and RQ4b) how does multiple entry in the same examination series influence outcomes? As discussed previously, multiple entry is likely to impact on outcomes due to a range of factors including the way in which students are taught and prepared for assessments, student motivation, and the presence of measurement error in assessment. The latter will always impact upon outcomes, even in a system of single examination entry, yet multiple entry is likely to exacerbate the effect of measurement error at the cohort level since students receive their highest grade across multiple attempts as their final grade. Conversely, in a system of single examination entry students that under- or over-achieve relative to their 'true' score will likely cancel one another out (see earlier discussion in this chapter).

Exploring the impact of EMEE on outcomes is challenging since it is difficult to disentangle the range of factors that might influence outcomes in the presence of EMEE (see earlier discussion in this chapter). However, a number of comparisons were undertaken to provide some insight into the effect of EMEE on outcomes. First, all students' entry attempts and each student's best grade were compared for students completing Year 11 in the 2012 – 2013 academic year. Second, the difference between a student's best grade and their average grade were compared, according to the number of entry attempts each student made. Finally, the first analysis was repeated for those students that entered twice in the same examination series.

All vs. best grades

The first analysis compared the difference between all students' entry attempts and their best grade, shown in Table 21 as the cumulative percentage outcomes at each of the key GCSE grades (see Chapter 1). As shown, these outcomes differ fairly significantly, particularly at grade C, where there is a 16.20% difference in outcomes. There are also differences at the higher grades of A* and A, yet these are smaller. The larger differences at grade C are perhaps not surprising considering that this grade was a key threshold for school performance measures and for influencing individual student's future life chances

(see Chapter 1). Consequently, approaches such as EMEE were particularly relevant for students around this threshold (see Chapters 4 and 5 and earlier analyses in this chapter).

The scale of these differences might lead one to conclude that multiple entry has had a large effect on outcomes, but this cannot be assumed from the data presented here. While some of the differences in outcomes might be due to measurement error increasing a student's best grade relative to their 'true score', the differences may also represent legitimate improvements in performance between examination series, for example due to a maturation effect (although one might argue that students should not have entered early if they were not ready; see Chapter 2). The differences may also be due to differences in student motivation or the extent to which students are prepared for the assessments. Evidence from Chapter 5 suggests that some students were entered early for assessments with insufficient preparation (for a variety of reasons). As such, this is likely to have influenced the grades that students achieved on some of their entry attempts and had a deflationary effect on outcomes when all entry attempts are considered. Thus, the comparisons in Table 21 are likely to over-emphasise how EMEE (in itself) might impact upon outcomes.

Table 21. *Cumulative percentage outcomes of all vs. best entry attempts*

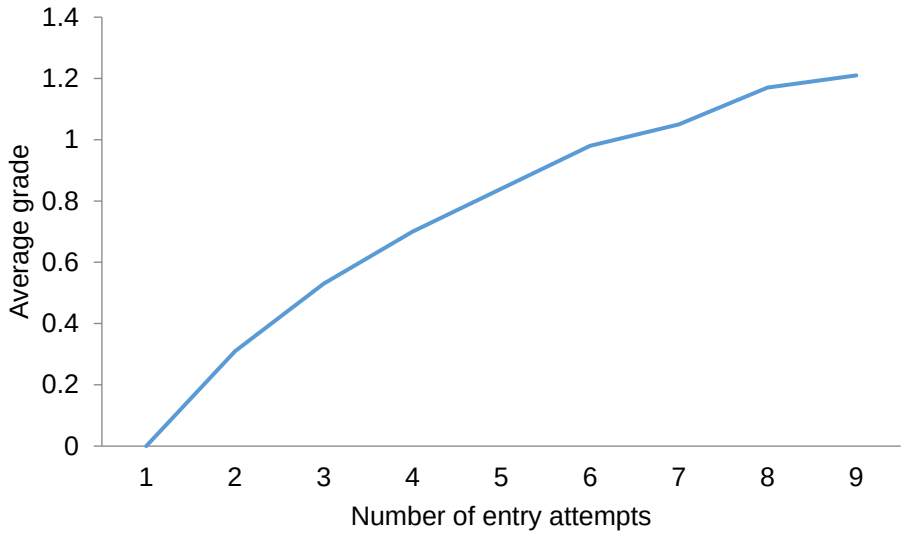
Entry attempts	A*	A	C	F	Number of entries
All entries	4.81	13.57	57.45	92.76	1,055,890
Best entry	8.41	21.55	73.65	95.11	595,240

Average vs. best grade

The second comparison explored the difference between a student's best grade and their average grade, according to the number of entry attempts made. To do this, students' grades were converted to numbers (A*=8, A=7, B=6, C=5, D=4, E=3, F=2, G=1, U=0) and the average difference between each student's best grade and their average grade plotted against the number of entry attempts each student made (see Figure 16). This revealed

that, as the number of entry attempts increased, the difference between a student's best grade and their average grade also tended to increase[93].

Considering this, one may again conclude that the greater number of entry attempts has increased outcomes. However, there are a number of other factors that could have contributed to this. For example, Figure 14 showed that the weaker students (who are likely to perform poorer) entered on a greater number of occasions than the stronger students. These students are likely to have a lower baseline grade and therefore have greater room for improvement, potentially increasing the difference between their average and best grade. In addition, student motivation, teaching, and when a student entered might also be influential (see Chapter 5). As the number of entry attempts increases, student motivation may decrease, meaning that performance is affected and a student's average grade is deflated. Furthermore, students may have been differentially prepared for each of their examination attempts, which may have influenced performance: if students were under-prepared (for any reason) their performance may have been lower, which could again deflate their average grade.



[93] Data for students with more than nine entry attempts was excluded, since there were fewer than 100 students in each case (see Table 14).

Figure 16. *Mean difference between highest and average grades by number of entry attempts*

Double entry in the same examination series – all vs. best grades

The final analysis aimed to begin to disentangle some of the factors that might influence the findings described above, by considering outcomes for students certificating twice in the same examination series (see also Chapter 2). If it is assumed that students entering twice in the same examination series will have similar levels of preparedness and motivation for each assessment, then theoretically at least, their grades should be the same (if it is also assumed that the standards of the qualifications are the same – see Chapters 1 and 8). However, the presence of measurement error in assessment means that a student might achieve a different grade on each assessment by chance – not due to differences in their ability or levels of preparedness – but due to the random error inherent in measurement (Wiliam, 2001)

The analysis described here focused on students that entered twice in the same examination series, regardless of the total number of entries that they made. As such, the analysis included students that entered twice where these entries were in the same examination series, and students that entered on up to 12 occasions where two of these entries were in the same examination series. The analysis excludes any students that entered more than twice in the same series.

The analysis was also restricted to students entering twice in the same summer series, since it is not possible to identify which series (November, January or March) a winter entry refers to (see earlier discussion in this chapter). Furthermore, students were only included if they had sat two GCSE mathematics qualifications: entries to international GCSEs were excluded since the nature of the assessment is likely to differ, suggesting that students might have been better prepared for one of the two specifications. This left a total of 17,356 students.

Table 22 shows a cross-tabulation of grades for students entering twice in the same summer examination series (note it is not possible to determine which assessment was sat first). This shows that just over half of students achieved the same grade in both assessments (51.12%), meaning that a significant number of students achieved a different grade on each assessment. Where this was the case, students typically achieved a grade difference of one (43.14%), although there were also instances where the differences were greater. For example, 5.74% of students achieved results that were two or more grades different, and the largest difference was five grades. The more extreme cases might reflect a lack of motivation where a student engaged with one assessment more than the other, or students might have been partially absent for one of the assessments in a qualification and thus performed poorer (partial absentees are not identified in the NDP; see earlier discussion in this chapter). Generally though, these findings reflect previous analyses (see Chapter 2): both Vidal Rodeiro (2014) and Keeling (2009) found that around half of students entering two assessments in the same examination series achieved the same grade in each.

It is also apparent from Table 22 that double entry students tended to be those around the grade C/D borderline. To some extent, this is likely to reflect the spread of students across GCSE grades, with a greater proportion concentrated around the middle grades. However, it might also reflect some of the evidence presented thus far in this thesis (see Chapters 4 and 5 and the earlier analyses in this chapter). For example, schools reported giving grade C borderline students as many entry attempts as possible (see Chapter 5), and the use of multiple entry in the same examination series would facilitate this. Again, this reflects previous findings (see Chapter 2).

Table 22. *Cross-tabulation of grades for double entry students*

		Grade 2									
		A*	A	B	C	D	E	F	G	U	Total
Grade 1	A*	113	78	1	3	2	0	0	0	0	197
	A	92	613	308	12	4	5	0	0	0	1,034
	B	5	289	1,533	763	22	42	7	0	0	2,661
	C	1	8	713	2,837	1,595	104	48	5	2	5,313
	D	0	7	23	1,429	2,070	664	105	12	16	4,326
	E	0	12	34	127	644	862	239	33	34	1,985
	F	0	0	18	60	86	283	428	114	31	1,020
	G	0	0	0	6	12	38	112	275	84	527
	U	0	0	0	2	11	26	32	81	141	293
	Total	211	1,007	2,630	5,239	4,446	2,024	971	520	308	17,356

The differences in grades illustrated in Table 22 were explored further by considering outcomes when comparing a student's best entry attempt[94] against both of their entry attempts (see Table 23). This comparison aimed to gauge the extent to which multiple entry in the same examination series – when levels of preparedness for the assessments should be similar – might impact upon outcomes. As shown in Table 23, the difference in outcomes between all entry attempts and a student's best grade were minor at grades A*, A, and F, yet a large difference was observed at grade C. Indeed, when using a student's best grade outcomes were over 10% higher. The greater difference at grade C is likely to be related to the propensity for students around this threshold to be entered twice in the same series (see Chapter 5 and earlier discussions in this chapter).

These findings suggest that multiple entry in the same examination series can have a large impact on outcomes. Furthermore, these differences are less likely to be attributable to levels of preparedness or motivation, since the assessments were taken in the same

[94] A student's best grade was defined as the best grade across these two entry attempts, though it is noted that they could have achieved a higher grade in another examination series.

examination series. This raises concerns around the extent to which the grades that students achieve reflect their 'true score' if they are achieved through methods such as double entry (see Chapters 7 and 8 for further discussion).

Table 23. *Cumulative percentage outcomes of both vs. best entry attempts*

Entry attempts	A*	A	C	F	Number of entries
Both entries	1.18	7.06	52.70	95.25	34,712
Best entry	1.70	8.95	62.94	96.65	17,356

Discussion

The analyses in this chapter have explored the possible implications of EMEE for outcomes in GCSE mathematics. Despite the caveats (that have previously been discussed), the findings suggest that EMEE is likely to have implications for examination outcomes at the cohort level. More specifically, that EMEE is likely to increase outcomes. This could be due to a number of factors such as the extent to which students are prepared for assessments, their motivation and engagement, and the presence of measurement error in assessment. Disentangling the effects of these factors is not possible given the data available here, but the analyses comparing outcomes when students enter on multiple occasions in the same examination series provides some insight. These analyses suggested that when factors such as test preparedness and motivation are assumed to be relatively stable (as far as can be assumed), EMEE has an effect on outcomes, particularly at grade C.

The findings in this chapter provide insight into the third and fourth research questions addressed in this thesis. In addition, the results are useful in other respects – in particular, the outcomes reported in Table 21. The data presented here provide an indication of the final outcomes achieved at each grade for GCSE mathematics for the cohort of students who completed Year 11 in summer 2013 (when considering their best grade; see Table 21). This might not seem novel in itself, but it should be considered that this data is not

routinely available in the public domain.[95] Although national GCSE outcomes are reported each August (see Chapter 1 and JCQ, 2016, for example), these only refer to entries from the summer examination series. This means that in the presence of EMEE these figures do not provide an accurate reflection of how the cohort of students each year has performed, since some students may have achieved their highest grade in an earlier examination series, while others may not even enter the summer examination series so will not contribute to the published results. This would have particularly been the case when EMEE was increasing, since GCSE mathematics was available in multiple examination series each year (see Chapter 1).

This raises some interesting points around how national results are reported and interpreted within the public domain. Although this is not the specific focus of this thesis, it is considered briefly here. In the presence of EMEE, the outcomes reported on results day each August will likely differ significantly to those that would be calculated when considering each student's best grade across all examination series – an effect illustrated in this chapter. Despite this, national results are typically subject to considerable media attention (see Chapter 1) and are often (rightly or wrongly) used as a metric for judging standards in education. Indeed, small changes in outcomes between successive years can prompt considerable debate amongst stakeholders: a rise in outcomes can lead to debate around whether standards are being lowered or whether students are performing better (Murphy, 2004; Stringer, 2012), while a fall in outcomes can lead to accusations of 'meddling' by the regulator and politicians, as in the case of GCSE English in summer 2012 (Sellgren, 2012; Vasagar, 2012). It is therefore of concern that, in the presence of changing entry patterns brought about by EMEE, such comparisons are not likely to be valid, since the population of students certificating in each summer series is constantly in

[95] Note that the DfE produce statistical reports containing results data each year. However, no overall measure of outcomes across all students and school types in GCSE mathematics could be identified in these.

flux. This is particularly likely if schools adapt their entry approaches on a regular basis, a practice reported by some respondents in Chapter 5, and is of concern if national outcomes are used as a metric for judging standards in education.

Summary of Chapter 6

To summarise, this chapter has explored the prevalence of EMEE, the type of students that enter on single or multiple occasions, and the potential impact of EMEE on outcomes in GCSE mathematics, focusing on students completing Year 11 in summer 2013. In many respects, the focus on GCSE mathematics is likely to represent the most extreme case of EMEE, since this subject is thought to lend itself more readily to EMEE than most other subjects (see Chapter 1). Therefore, it is unlikely that the findings reported here would be generalisable to other subjects, including GCSE English, despite the mandatory inclusion of this subject in school performance measures.

The analyses focused on the 2012 – 2013 academic year. Here, around half of the cohort entered on more than one occasion prior to their completing Year 11. Analyses of students' outcomes and prior attainment data suggested that the choice of entry route was not arbitrary: single entrants tended to be the higher ability students who performed well, but as the number of entries increased, ability and outcomes declined.

EMEE was shown to impact upon outcomes when comparing a student's best grade against all of their entry attempts, particularly at grade C. Furthermore, there was evidence that as the number of entry attempts increased, the difference between a student's best grade and their average grade also increased. While this suggests an effect of EMEE on outcomes, it is not possible to disentangle all the factors that may have contributed to these differences – in particular, to ascertain whether outcomes are legitimately different or whether they are increased due to the presence of measurement error in assessment. Thus, the final analyses explored the impact of double entry on outcomes when students certificated twice in the same examination series. This showed a

large effect on outcomes at grade C and a smaller effect at the other key GCSE grades (A*, A and F). This effect is arguably less likely to be influenced by levels of test preparedness or motivation since the assessments were sat in the same examination series, and is explored further in Chapter 7, where simulations are used to model the extent to which multiple entry might impact on outcomes when factors such as preparedness and motivation are ignored.

Chapter 7 – Simulating the potential impact on outcomes

Overview

This chapter explores the potential effects of EMEE on GCSE outcomes at the cohort level using simulations. Specifically, it illustrates how outcomes can change as the proportion of students sitting assessments on multiple occasions varies. The simulations complement the empirical analyses in Chapter 6, where trends in EMEE and the effect on overall cohort outcomes were explored using the NPD. The findings from Chapters 6 and 7 are discussed in light of the implications for maintaining examination standards in Chapter 8. Thus, this chapter focuses on the simulation study, rather than the possible implications for examination standards.

Introduction

The analyses in Chapter 6 focused on the prevalence of multiple entry in GCSE mathematics and the potential impact on examination outcomes in the 2012 – 2013 academic year. The latter part of these analyses revealed a large difference in cohort outcomes, particularly at grade C, when comparing a student's best grade with all students' grades or a student's average grade in the presence of EMEE. Furthermore, there was a large difference in outcomes when comparing a student's best grade with all students' grades for those who certificated twice in the same examination series. The potential reasons for this and the implications have been discussed: for example, the differences could be due to student level variables such as exam preparedness or motivation, or due to the unreliability that is inherent in assessment, meaning that students might achieve a different grade on each testing occasion (see Chapter 6).

Disentangling these factors in the analyses presented in Chapter 6 was not feasible, since quantifying the extent to which students were prepared or motivated for different assessments was not possible given the data available. However, the final analysis that

compared all students' grades with a student's best grade for those that certificated twice in the same examination series provided some insight into the relative influence of different factors, since students certificating multiple times in the same series would be more likely to have similar levels of preparedness and motivation for each assessment (though the latter cannot be assured). Thus, it might be argued that this difference in outcomes, which was large at grade C (10.24%), is more likely to be influenced by the error present in assessment than the characteristics of the individual students.

The following chapter builds on the findings in Chapter 6 by considering the potential influence of EMEE on outcomes using simulations. Simulations are used since they allow the extent of multiple entry to be varied – i.e. the proportion of students entering on multiple occasions – and the impact on outcomes illustrated. Furthermore, they allow us to explore the possible effects of EMEE on outcomes when student level factors are assumed to be constant. These analyses therefore contribute to RQ4: what are the implications of EMEE for examination outcomes in GCSE mathematics assessments? In doing so, they address the following subsidiary research questions, continuing on from those outlined in Chapter 6:

> RQ4c) How does multiple entry impact upon cohort outcomes as the proportion of students entering on multiple occasions varies?
>
> RQ4d) More specifically, what are the differences in cohort outcomes when comparing all students' grades with each student's best grade, as the proportion of students entering on multiple occasions varies?

When considering the findings it is worth bearing in mind that the simulations are modelling the extent to which outcomes change in the presence of EMEE, and that any findings are illustrative, since they do not account for any student level variables that are present in the live data – factors such as exam preparedness and motivation. Thus, the simulations do not attempt to replicate the situation observed in practice (see Chapter 6). Rather, they aim to provide an illustration of how multiple entry can influence outcomes as

the proportion of students entering on multiple occasions varies. The implications of this for maintaining examination standards is discussed in Chapter 8.

Methodology

To illustrate the impact of EMEE on outcomes, the analyses in this chapter reflect those reported in Chapter 6 that explored the difference in outcomes between a student's best grade and all students' grades in the presence of EMEE. However, instead of being based on live data, the outcomes are simulated. Thus, a cohort of students is generated who are deemed to be either single or multiple entrants. This is achieved by assuming that there are two assessments available to students, that a cohort of students is sitting each test, and that a certain proportion of these students are sitting both tests. Those students sitting just one test are single entrants, while those sitting both tests are multiple entrants. Although Chapter 6 revealed a wide variety of entry approaches (i.e. some students entered on up to 12 occasions), the analyses reported here focus upon double entry only, since double entry was by far the most prevalent approach when students entered more than once (see Chapter 6). Furthermore, the simulations are intended to be illustrative only.

For each student, test scores are simulated (on one or both tests, as appropriate) and these scores are converted to grades. This allows cumulative percentage outcomes to be generated:

- on each test separately (including those students that sat one test or both tests),
- aggregated across both tests (including every grade such that multiple entrants are counted twice), and
- using a student's best grade (on either test).

The outcome of interest is the difference in cumulative percentage outcomes when considering all entry attempts and a student's best entry attempt.

As discussed in Chapter 6, double entry could stem from students entering two different specifications in the same examination series, or students entering the same or different specifications in different examination series. However, in the context of these simulations, the distinction between students entering multiple specifications or with multiple awarding bodies is not significant, since the simulations aim to illustrate the effect of varying levels of multiple entry on outcomes and not account for differences in preparedness or maturity between different assessment attempts. Indeed, it is assumed that our simulated students are equally well prepared and motivated for each test. Given this, the two tests could be considered as parallel forms of the same test, rather than two specifications from the same or different awarding bodies.

A detailed description of the methodology is provided below but, in brief, the simulations;

i) Generate a set of students sitting one test or both tests, each with a true score – this gives a profile of students and their relative ability;
ii) Generate a conditional sum score probability distribution (cssd) around each true score – this gives the probability of a student achieving any score on the test given their true score;
iii) Select an observed score for students on one test (single entrant) or both tests (multiple entrant) given their true score and the associated cssd;
iv) Grade the students on each test;
v) Calculate the difference in cumulative percentage outcomes when comparing all students' grades and a student's best grade;
vi) Manipulate the proportion of students entering on multiple occasions and calculate the difference in outcomes.

All analyses were conducted using R (R Development Core Team, 2014). The following section provides details of each stage of the methodology.

i) Generating the true scores on each test

To generate a set of students sitting each test, a distribution of true scores available on each test was generated, from which students sitting each assessment were selected (see Figure 17 for an example). This distribution of true scores was the same for each test and was based on the beta-binomial distribution,[96] using a mean and standard deviation that is specified for each test. The distributions were based on the number of items specified (80 dichotomous items) and what could be considered 'ideal' values for the mean and standard deviation for assessments (the mean was set as 50% of the maximum mark and the standard deviation as 16.7% of the maximum mark). These values are (informally) considered the 'ideal' values for a GCSE assessment since they can facilitate the setting of well-spaced grade boundaries by achieving a high enough spread of students across the mark range. Such outcomes would also suggest that the difficulty of the test is appropriate relative to the ability of the candidature.

An 80 mark test was modelled since this allowed a level of reliability for the tests that is reasonable to assume in practice (test-retest correlation of 0.9). Although 80 marks is considerably less marks than would be used for a whole GCSE specification (and is more akin to an individual GCSE examination), changing the number of marks would narrow or widen the distribution of the cssd around each true score to suggest a level of reliability that might not be feasible in practice. In other words, given their true score, students would be more likely to achieve a score that would not be feasible in practice – i.e. it would be much closer or further from their true score than is realistic to expect.

[96] The beta-binomial distribution is used since this distribution has discrete end points; a normal distribution would not be bounded between the maximum mark and zero, creating difficulties at the edges of the distributions.

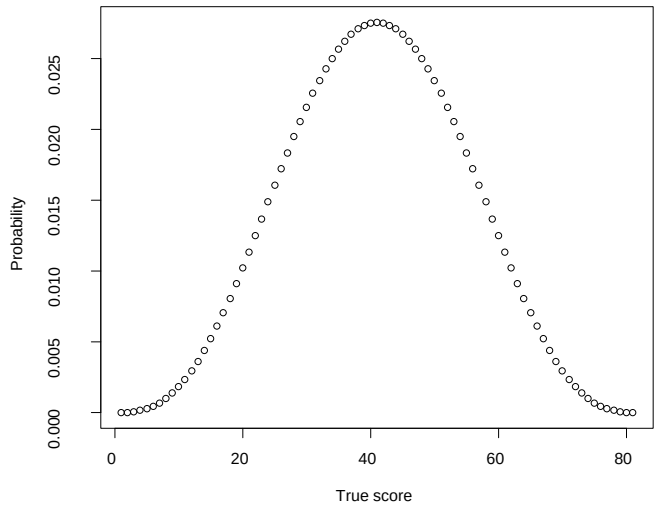


Figure 17. *Distribution of true scores on test A from which students are selected*

ii) Generating a conditional sum score distribution (cssd) for each true score

A conditional sum score probability distribution (cssd) was then generated for each true score. This distribution gave the probability of a student with a given true score achieving any score on the test, with the probability distributed around each true score. The cssds were generated using Monte Carlo simulations based on the Rasch model (Rasch, 1960), using the following procedure:

- An initial sample of student thetas (i.e. representing the ability of the students) and item betas (i.e. representing the difficulty of each item) were drawn from a standardised normal distribution. In this case, 10,000 students were used and an 80 item test of dichotomous items, giving 10,000 thetas (student abilities) and 80 betas (item difficulties).
- Using the Rasch model, for each item and each value of theta, the probability of a student getting a mark for each item was determined (i.e. the probability of getting the item correct). This gives a matrix containing probabilities of success for every student on every item.

- The score for each student on each item was estimated, given their probability of getting the item correct – this generates a set of simulated marks for each student on each item.
- The Rasch model was fitted to this data to generate item difficulties and theta's for the simulated assessment – these values were used rather than the original values sampled in i) since they fit the parameters specified, i.e. there were only item-1 possible theta's rather than the infinite number that could have been drawn from the original distribution.
- The Wingersky-Lord recursive algorithm (Lord and Wingersky, 1984) was used to determine for each theta, the probability of achieving any score on the test. This algorithm calculates the probability of someone with a given theta achieving all possible marks on the test via each possible route, i.e. a mark of 80 can only be achieved via one route (scoring one on every item), while a mark of one can be achieved in 80 different ways (scoring one on any of the 80 items). These probabilities give the probability that a student with a given ability (theta) will achieve any mark on the test (i.e. the cssd).

iii) Generating the students

The size of the cohort sitting each test (100,000) and the number of students entering on multiple occasions (see below) was then specified, allowing each students' true and observed scores to be generated. For students entering one test (i.e. entering once) their true score was sampled from the distribution of true scores as generated above for each test, resulting in a distribution of true scores for the students entering each test (see Figure 18 for an example). The observed scores for these students were then sampled from the cssd associated with each true score, generating a set of observed scores on each test for students entering once. A student's true and observed score was defined according to classical test theory (or true score theory; see Chapter 6): that is, each student's true score on the test was dictated by their true ability and, in conjunction with

an error term, comprised their observed score. The true score can therefore be considered as '*an expected (average) value of the test score, averaged over those factors classified as measurement error*' (Livingston and Lewis, 1995) – or the score that a student would theoretically achieve if all error (i.e. factors not relating to a student's ability) could be removed (Gardner, 2013).

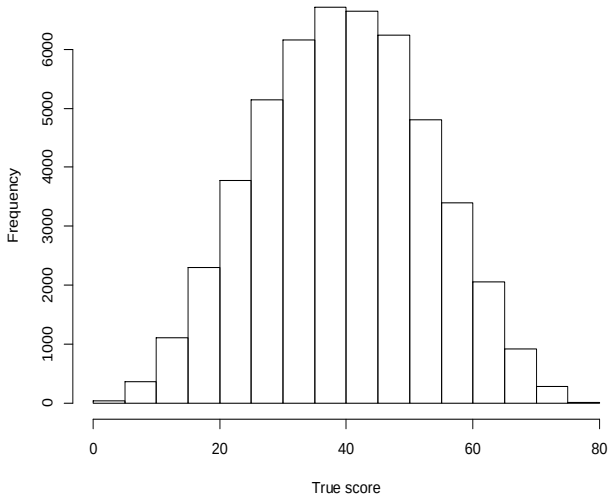


Figure 18. *Distribution of students' true scores for students entering test A*

For students entering both tests (i.e. multiple entrants), one true score was sampled from the distribution of true scores, since it was assumed that each student had the same true score for each test (see Chapter 6). The observed scores for these students were then generated in the same way as for students entering a single test, but for both tests (i.e. they have an observed score for each test).

iv) Grading the students

Once students' observed scores had been generated, these were converted to grades to allow a consideration of the impact of multiple entry on outcomes. This is done in much the same way that would occur in practice during awarding (see Chapter 1), as follows: i) a prediction was specified giving the expected outcomes for each test at each grade, ii) a

cumulative percentage distribution of observed scores for students sitting each test was generated, and iii) grade boundaries were set at the key grades (A, C and F; see Chapter 1) to ensure that the outcomes achieved by the students matched the predicted outcome as closely as possible. The remaining grade boundaries were then set arithmetically – for example, grade B was set halfway between grades A and C.

The predictions used in the simulations were based on the values that are observed for a typical cohort of students sitting GCSE mathematics in a summer examination series. By typical, this means that the predictions refer to 16-year-olds certificating at the end of Year 11, the cohort that is usually used to generate predictions in a summer examination series. This approach allows a spread of students across the grade range that would likely be observed in practice. The same predictions were specified for each test, thereby assuming that the prior attainment profile of the two cohorts, and the standards of the two qualifications, were the same.

The procedure described here sets ‘subject’ level grade boundaries that are used to grade students and calculate grade outcomes for a cohort of students. However, at the time that EMEE was increasing this was not how grades were assigned in practice, since most GCSE assessments had a modular structure (but all modules were sat concurrently from June 2014) and the modules were aggregated via the uniform mark scale (see Chapter 1). This method of aggregation was necessary when GCSEs were modular, since students could sit different modules in different examination series and the grade boundaries changed each series. Thus, raw scores on each assessment were converted to uniform marks to ensure comparability between students sitting assessments in different examination series – for example, a grade A in one series might require a different number of raw marks to a grade A in another series, yet the number of UMS marks for a grade A would always remain the same. From June 2014 onwards, students were required to sit each module in the same examination series (i.e. linear assessment), but since the structure of the qualifications had not changed (only the requirement to sit all

modules at once had been introduced), the aggregation method remained until the introduction of new GCSEs first assessed in some subjects in summer 2017. These new qualifications have been designed with a linear structure in mind and will be aggregated by an addition method (i.e. adding the marks on each component together), since students must sit all the assessments in the same examination series and there is no requirement to convert marks to a uniform scale. Thus, the method described here is more akin to that which is used for the new GCSEs (first assessed in 2017 for English language, English literature and mathematics), rather than the qualifications that were available when the work in this thesis was conducted.

v) Calculating grade outcomes for 'all' and 'best' entry attempts

Once students had been graded, the overall grade outcomes for the cohorts were calculated in the presence of single and multiple entry, with the differences between these suggesting the influence of multiple entry on outcomes. Thus, cumulative grade distributions were generated for each test, the combined tests (without taking into account multiple entrants, i.e. some students will have multiple results), and combined across tests using a student's best grade (as would be the case in practice). The differences between the aggregated combined outcomes across all grades and those using a student's best grade suggest the difference due to multiple entry.

vi) Manipulating the proportion of multiple entrants

Using simulations allows the proportion of multiple entrants to be modified and the effect on outcomes illustrated. Thus, the percentage of multiple entrants was specified at 5% intervals between 5% and 95% and the modelling repeated. The percentage of multiple entrants refers to the number of students that were deemed to be multiple entrants as a proportion of the total entry combined across both tests. For example, if there was a cohort of 100,000 students sitting each test and 5% of students for each test were multiple entrants (i.e. 5,000 students from *each* test were multiple entrants so there were 10,000 overall sitting both tests), then the total cohort would be 200,000 students and 10,000 of

these would be multiple entrants (i.e. sitting both tests). Each simulation was run 100 times and the average of the differences between all grades and best grades taken.

Results

The results summarised in Figure 19 show the difference in outcomes when considering all entry attempts and a student's best entry attempt for the overall cohort, plotted against the percentage of students that entered on multiple occasions (i.e. that entered both tests; see Appendix J for the actual figures). The results are presented for each key GCSE grade (see Chapter 1) – A, C and F – and the difference shown is the average difference across the 100 simulations (see Appendix J for the results for all grades). An example output from one run of the simulations is also provided in Table 24. This example refers to 5% multiple entry and shows the cumulative percentage outcome at each grade for each test individually (n = 100,000 for each test), for all grades combined across the two tests (including both grades for multiple entrants; n = 210,000),[97] and for a student's best grade (n = 200,000). The variable of interest is the difference in outcomes between all grades and students' best grades.

Before considering the findings, there are two points worth noting. First, when discussing the results, the focus is on the overall trends rather than the specific numbers or how these compare to the figures in Chapter 6. The reasons for this are summarised in the discussion. Second, it is worth mentioning that prior to summarising the findings, consideration was given to the output and consistency of the simulations. This revealed that when the percentage of multiple entrants was the same, there were only minor fluctuations in the differences between all students' grades and students' best grades for each run of the simulation. Such minor fluctuations are not unexpected due to the random

[97] The sample is 210,00 as there are 95,000 single entrants on each test and 10,000 multiple entrants, who each have two grades.

sampling of students each time the simulations were run and are not considered problematic.

Table 24*. Example output from simulations with 5% multiple entry*

	A*	A	B	C	D	E	F	G	U
Assessment 1	7.34	19.70	37.89	59.04	75.72	88.51	95.41	98.80	7.34
Assessment 2	6.28	19.63	40.48	61.25	77.90	88.47	95.35	98.77	6.28
All grades	6.81	19.66	39.19	60.14	76.81	88.49	95.38	98.79	6.81
Best grade	6.92	19.89	39.48	60.47	77.07	88.65	95.46	98.83	6.92
Difference	0.11	0.22	0.30	0.32	0.26	0.16	0.09	0.04	0.11

Turning now to Figure 19, it is evident that as the proportion of multiple entries increase, the difference between all students' grades and each student's best grade also increases. In other words, the final outcomes that students achieve at the cohort level (denoted by their best grade) increase as the proportion of multiple entrants increase. This is not unexpected since as the number of entry attempts increases, so does the chance that students will achieve a grade higher than their 'true' score would suggest (see Millman, 1989; Wheadon, 2010). Thus, the more students that are entering on multiple occasions, then the greater the number of students who are likely to achieve a grade above their 'true' score. Multiple entry also reduces the chance of a false negative (Wheadon, 2010) – that is, a student achieving a score or grade lower than their true score would suggest (Cizek and Bunch, 2007). Therefore, students that were multiple entrants and performed poorly on one test were likely to achieve a higher score or grade on the other test, again increasing outcomes.

The results show that the greatest difference in outcomes (or the greatest increase in outcomes) as multiple entry increases is at grade C, followed by grade A then grade F. This is not surprising since the distribution of students across the ability range means that there are a greater proportion of students concentrated around the middle of the grade

range, where grade C is typically situated (in these simulation 60% of students were predicted to achieve a grade C). This effect would typically be observed in practice for large entry subjects such as GCSE mathematics, since the ability of the cohort is spread across the ability range so that the greatest proportion of students are in the middle.

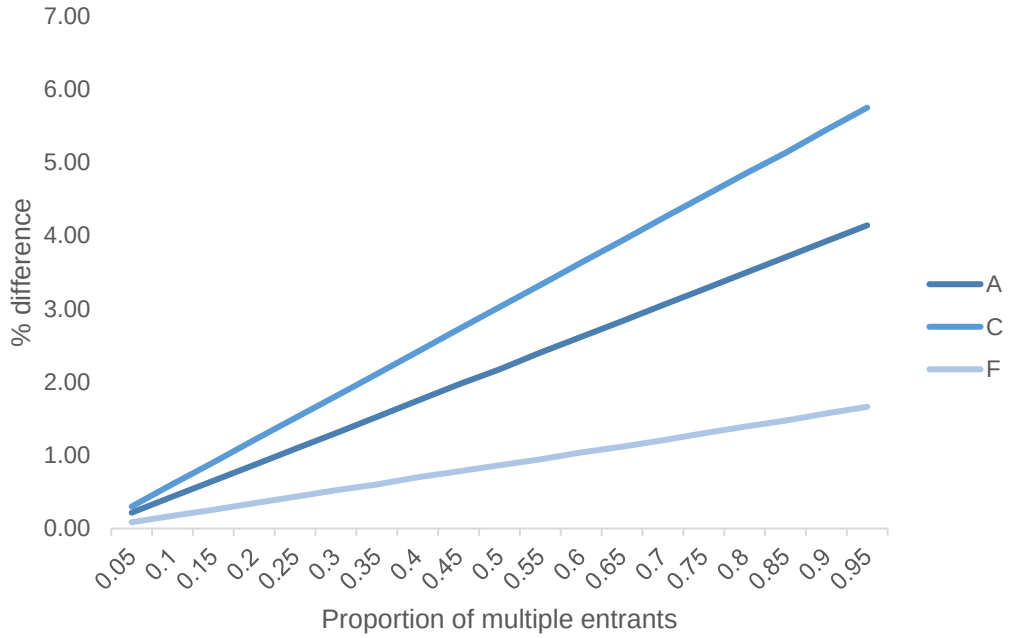


Figure 19. *Differences between all grades and students' best grade by proportion of multiple entrants*

To summarise these findings, Figure 20 illustrates the outcomes for the whole cohort of students as the proportion of multiple entry changes, focusing on grade C (the effects are similar but smaller at the other key grades). Thus, while the prior ability of the cohort is similar each time and the standard of the qualifications is assumed to be the same, the fact that more students are entering on multiple occasions increases the final outcomes. Indeed, the predictions for both tests were for 60% of students to achieve a grade C, which is exceeded to a greater extent as the proportion of multiple entrants increase. This illustrates how outcomes can differ due to the entry approach that is used, rather than due to differences in the ability of the students or the standard of the qualifications, and raises

concerns around the implications for maintaining examination standards and interpreting outcomes. These issues are discussed in Chapter 8.

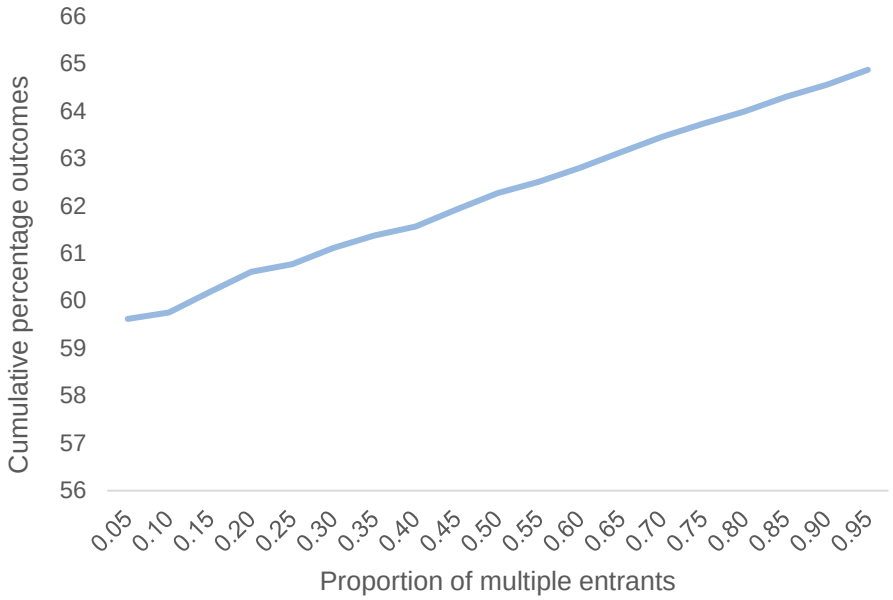


Figure 20. *Cumulative % outcomes (best grade) and the proportion of multiple entrants – grade C*

Discussion

This simulation has illustrated the potential effect of multiple entry on examination outcomes, by manipulating the proportion of students entering on a single or multiple occasion. The results show that, as the proportion of students entering on multiple occasions increase, the difference in outcomes between all students' grades and their best grade also increases. Moreover, as the proportion of multiple entrants increase, examination outcomes for the overall cohort of students also rise, particularly around the grade C threshold. These findings suggest that when the ability of the student cohort and the standard of the qualifications remain constant (as simulated here), the practice of multiple entry itself can increase the outcomes achieved when a student's best grade counts. Thus, it is possible that any differences in the final outcomes (in terms of students'

best grades) between two cohorts could reflect the way in which the students were entered for the assessments, rather than the ability of the cohort or the standard of the qualifications. This will have implications for how grades and outcomes can or should be interpreted, and the process via which standards are maintained – issues that are discussed in Chapter 8.

The comparisons in this chapter are similar to those reported in Chapter 6 that explored the difference between all students' grades and their best grades in the presence of multiple entry using empirical data from the 2012 – 2013 academic year, yet the outcomes from the live data and the simulations differ fairly markedly. For example, in the empirical data there was a 16.20% difference in outcomes between all attempts and a student's best attempt when multiple entry was around 50% of the cohort (see Chapter 6), yet the simulations suggest a difference of 3.02% for the same proportion of multiple entrants. This is not considered surprising, or indeed problematic, since there are several key differences between the two sets of analyses. In particular, the simulations do not aim to replicate the situation observed in practice or take account of the many factors that could influence examination outcomes – for example, student motivation and preparedness. Rather, they aim to illustrate the potential effect of EMEE on examination outcomes due to the unreliability of assessment, as the proportion of multiple entrants change. Thus, there are limitations to the conclusions that can be drawn and the actual figures should only be considered as illustrative. Furthermore, there are several limitations and assumptions when using simulations, as outlined below.

Limitations of simulations

Although the simulations have illustrated the potential effect of multiple entry on outcomes, there are several limitations that require consideration. First, the simulations do not account for any student level variables – for example, student preparedness and motivation – that are likely to be influential in practice. The simulations therefore essentially assume that students have sat two parallel forms of the same test at the same

time, while in practice they are likely to sit different specifications with different awarding bodies in different examination series. Thus, while the simulations have allowed the proportion of multiple entrants to be varied and the potential effect on outcomes explored, the outcomes should only be considered as illustrative.

Second, multiple entrants were randomly selected from across the ability range. As discussed in Chapters 5 and 6, multiple entry is typically concentrated amongst the C/D borderline students – i.e. the middle ability students. However, in the simulations multiple entrants were drawn from across the whole population. While multiple entrants would still have been concentrated in the middle of the ability range due to the distribution of true scores that the students were selected from (see earlier discussion in this chapter), in practice, multiple entrants are likely to be much densely concentrated in the middle of the ability distribution. It is likely that had the multiple entrants been sampled specifically from the middle ability students, then the effect on outcomes at grade C/D would have been greater. Thus, the simulations likely under-estimate the effects of assessment unreliability and multiple entry on outcomes at this grade.

Finally, the analyses reported here have focused on double entry. However, as revealed in Chapters 5 and 6, students (and schools) adopted a wider variety of entry strategies than this, with some students entering on up to 12 occasions (see Chapter 6). As the number of entry attempts increase, so does the likelihood that students will achieve above the grade suggested by their ‘true’ score (see Millman, 1989; Wheadon, 2010). Thus, if multiple entry was simulated where students entered on a greater number of occasions, then the outcomes would likely differ to a greater extent than observed here.

Assumptions

In addition to these limitations, a number of assumptions have been made in the analyses. First, it has been assumed that students entering both tests have the same ‘true’ score for each test yet, in practice, one could debate whether this assumption would hold (see

Chapter 6), even if students were entering the assessments at the same time with the same levels of preparedness and motivation. This is particularly debatable if students were entering different specifications (e.g. modular or linear) or with different awarding bodies. Arguably, the assumption that the 'true' score would be the same is more likely to hold for GCSE mathematics than other subjects, since the content of the specifications is broadly comparable (a feature that has facilitated EMEE across awarding bodies and specifications – see Chapter 5). Furthermore, if the 'true' scores were not the same, then they would be correlated to a certain extent, since it would be anticipated that students would perform similarly on the two tests (i.e. a strong student would be expected to perform well on multiple tests, while a weak student would be expected to perform poorly on multiple tests). Thus, in the context of these simulations, this assumption seems reasonable, particularly if the tests are considered to be parallel forms of the same test (see Livingston and Lewis, 1995). However, if the tests were considered to be different specifications with different awarding bodies, then the assumption might not hold and the 'true' scores might therefore be better defined as highly correlated.

A second assumption is that student performance is approximately normally distributed on each test and that the predicted outcomes are spread across the grade range. While the latter is typically true in GCSE mathematics (the large cohort typically has a wide spread of ability), student performance on a test is not always normally distributed. If the mark distributions were skewed and students were concentrated in one part of the mark distribution, then this could narrow the grade boundaries and increase the chance of students being misclassified (Wheadon and Stockford, 2010) – i.e. a student being awarded a grade that does not reflect their true ability. In such situations, multiple entry could have a greater impact on outcomes, since the chance of a student falling on the other side of the grade boundary to that which would be associated with their 'true' score would be greater.

Summary of Chapter 7

In summary, this chapter has modelled the possible effects of EMEE on outcomes using simulations. While there are a number of limitations and assumptions to this approach, the analyses have illustrated that as the proportion of students entering on multiple occasions increase, outcomes also increase. This raises questions around how grades can be interpreted and how standards should be maintained in the presence of EMEE. The following chapter discusses these issues by considering the implications of EMEE for maintaining examination standards, including a consideration of the theoretical literature relating to how examination standards are defined.

Chapter 8 – Examination standards: the implications of early and multiple examination entry

Overview

This chapter draws together the findings from the previous chapters to consider how shifts in entries can have implications for maintaining examination standards. To understand these issues requires a reflection on how examination standards are maintained in practice, and a consideration of the theoretical literature relating to how examination standards are defined. The chapter is therefore arranged in three sections: the first section discusses how examination standards are defined, in relation to the comparable outcomes approach used to maintain standards in England (see also Chapter 1). The second section provides the context for later discussions by summarising the changes in entries to GCSE mathematics between 2007 and 2014 (focusing on EMEE), and the factors and policies that contributed to these changes (see also Chapters 1 and 2). The third section draws the first two sections together by discussing the implications of EMEE for maintaining examination standards – considering both operational and theoretical issues. These are both considered to be important – and related – issues. While one must ensure that effective arrangements are in place to maintain standards operationally, to do so requires a consideration of what one is trying to achieve. Essentially, to evaluate whether standards have been maintained requires a criterion, and this is informed by how standards have been defined.

What do ‘examination standards’ mean?

This chapter draws together the previous chapters of this thesis to consider the implications of EMEE for maintaining examination standards – both in an operational sense for how standards are maintained in live assessments, and theoretically for how standards are defined. While the definition of EMEE has previously been discussed, a

brief recap is provided here. As discussed in Chapter 1, students in England are expected to certificate (or complete) their GCSE qualifications once at the end of Year 11 (aged 16). Early entry has therefore been defined as instances where a student certificates throughout Year 10 or mid-way through Year 11 (regardless of whether they re-sit at a later date), and multiple entry as instances where a student certificates more than once before the end of Year 11 (regardless of whether this is with the same specification or awarding body or in the same or different examination series). These practices can be defined separately but are intrinsically linked: early entry is a pre-requisite for multiple entry, since it allows students multiple entry attempts before the end of Year 11. Thus, EMEE has been defined as instances where students certificate throughout Year 10 or midway through Year 11 and on multiple occasions before the end of Year 11.

The analyses presented in this thesis have shown how entry trends to GCSE mathematics changed between 2007 and 2013, particularly in relation to EMEE. The analyses in Chapter 4 showed that between 2007 and 2011 there was a sharp rise in the number of students entering GCSE mathematics (and English) early, a phenomenon reported elsewhere and known to have continued until 2013 (see Chapter 2 and Ofqual, 2014a). The analyses in this thesis have also shown how early entry was frequently coupled with multiple entry in GCSE mathematics: Chapter 5 presents qualitative evidence from individuals in schools describing concurrent early and multiple entry, while the analyses in Chapter 6 show widespread multiple entry in the 2012 – 2013 academic year. Thus, between 2007 and 2013 entry trends changed considerably: rather than students entering GCSE mathematics once at the end of Year 11, a wide range of entry approaches were used, particularly focused on EMEE.

The potential implications of these approaches for GCSE outcomes were explored in Chapters 6 and 7. This included a consideration of how EMEE might influence outcomes at the cohort level. Ultimately, this issue became obsolete due to changes in entries following the announcement in September 2013 that, henceforth, only a student's first

entry attempt would count towards school performance measures (see Chapter 1). Early (and therefore multiple) entry to GCSE mathematics effectively ceased following this announcement, and single examination entry at the end of Year 11 once again became the norm (e.g. see JCQ, 2014; 2015; 2016; Ofqual, 2014a). However, both the rise and subsequent fall in EMEE has had implications for examination standards, more specifically for ensuring comparability of standards, and that is the focus of this chapter.

This chapter therefore includes a consideration of how examination standards are set and maintained in practice for live assessments, given how examination standards are defined. The chapter is structured such that the first section discusses how examination standards are defined, drawing on the theoretical literature, since this provides the necessary context for considering the implications of EMEE for maintaining standards. The following section provides an overview of the factors and policies that contributed to the rise and fall of EMEE between 2007 and 2014. The final section draws the previous sections together by discussing the implications of EMEE for maintaining examination standards operationally (focusing on the rise and fall in EMEE between 2007 and 2014), given how examination standards have been defined.

Defining examination standards

Theoretical considerations of examination standards

The comparability of examination standards is an issue that has been debated for several decades (see Newton, 2007). Comparability is complex in England due to the presence of multiple providers offering qualifications in the same subject, yet it is key due to the uses that examination results are put (see Chapter 1). This issue has therefore generated a body of literature relating to both how standards should be defined and techniques for monitoring the comparability of examination standards (see Baird, 2007; Baird and Gray, 2016; Coe, 2010; Newton, 2005a, b; 2010a, b; Newton, Baird, Goldstein, Patrick, and Tymms, 2007). In this section we are concerned with the former: how examination

standards are defined. This is a complex and much debated issue, with commentators often discussing standards and comparability using contradictory definitions and terms (Baird and Gray, 2016; Coe, 2010; Newton, 2010a).

To consider how we might define examination standards in this thesis, we turn to the range of definitions that have been developed in the literature, focusing on the English assessment context. Throughout the debate on comparability, a number of authors have attempted to categorise definitions of standards, resulting in a range of typologies (see Newton, 2010a, for a summary). In his more recent work, Newton (2010a, b) sought to synthesise the existing literature and suggested three categories for defining examination standards: phenomenal, causal, and predictive. These categories each attempt to describe comparability in terms of what students that have achieved comparable results – for example, at a linked grade boundary – have in common, and are categories in the sense that alternative methods can be located within these (Newton, 2010a). In doing this, Newton (2010a, b) noted that methods and definitions should be distinguished from one another (see also Coe, 2010). Consequently, both statistical and judgemental techniques can be used to gather evidence within each category (Baird and Gray, 2016; Newton, 2010a).

Newton's (2010a) phenomenal perspective considers standards as comparable where the attainment (or outcomes of learning) are the same for students at key linked boundaries (i.e. at key grade boundaries; see Chapter 1). As described by Newton (2010a), students '*must be equally good at knowing, understanding and being able to do X, where X is a set of knowledge, skills and understanding that is common to both examination constructs'*. Newton (2010a) goes on to clarify that the phenomenal perspective does not necessarily construe comparability in terms of specific examination performance – rather, it is about attainment and the outcomes of learning. Given this definition, Newton (2010b) highlights the issue of comparability across different subjects areas – if comparability is considered

in terms of attainment, this creates challenges when considering comparability between maths and art, for example.

The second category, the causal perspective, suggests that standards are defined in terms of the causes of students' attainment – that is, comparability is defined by the extent to which students with similar background characteristics (e.g. prior attainment) achieve similar results. This means that even in instances where attainment can look quite different (e.g. between subjects), comparability can still be considered in terms of the causes of that attainment (Newton, 2010a). Whether one should be seeking comparability in such instances – for example between subjects – is another issue entirely though, and is not considered here (see Ofqual, 2015c, for an overview).

The final category, the predictive perspective, considers comparability as the extent to which students at linked boundary marks have the same chance of success in the future – that is, standards are defined by the future potential of students, given the outcomes that they achieved. This is described by Newton (2010b) as students having '*comparable prospects'* in the future.

The categories proposed by Newton (2010a, b) are not the only approach for defining standards, and are not without challenge. Coe (2010), for example, has suggested an alternative view – 'construct comparability'. Coe (2010) considers all definitions in terms of a variable that underlies attainment in the subject being assessed. Thus, comparisons between examinations can only be made if there is a common construct that they share – i.e. a common link between the examinations.

In addition to such alternative perspectives, more recently theoretical challenges to Newton's (2010a) definitions have emerged, particularly where the methods that are used for maintaining standards do not seem to align with any of the definitions (see Baird and Gray, 2016). However, since the purpose of this thesis is not to progress the debate on definitions of examination standards (but to consider the implications of EMEE for

examination standards), the categories outlined by Newton (2010a) are adopted. There are several reasons for this.

First, the categories proposed by Newton (2010a) provide a framework within which the meaning of standards in relation to EMEE can be considered. This would arguably not be possible in the same way under some alternative approaches (such as Coe, 2010), due to there being multiple possible approaches for linking standards that depend upon how the linking construct is defined (Newton, 2010a). Second, they provide a method of defining examination standards that is distinct from the methods used to maintain standards. This is important since the aim of this chapter is to consider the implications of EMEE for examination standards, given how they are defined theoretically. Finally, the definitions outlined by Newton (2010a) are situated within the assessment context that is the focus of this thesis – GCSE assessments in England (see Baird and Gray, 2016). Thus, the definition of standards adopted here provides the context within which one can evaluate the implications of EMEE for maintaining standards in practice. Given this, we now consider how the methods used for maintaining GCSE standards in England can be defined, using the categories outlined by Newton (2010a), before considering the implications of EMEE for examination standards.

Comparable outcomes and the meaning of examination standards

The process for maintaining standards for GCSE assessments in England centres on the comparable outcomes approach, first introduced in 2011 (see Ofqual, 2015b).[98] This approach prioritises statistical predictions and aims to ensure that once students' prior attainment is taken into account, students sitting assessments in different examination series are rewarded equally (see Chapter 1 for a full discussion). The comparable outcomes approach is rooted in earlier work by Cresswell (2003) and subscribes to the

[98] Note that exam boards also used statistical predictions prior to this (see Chapter 1).

'similar cohort adage' (Newton, 2011). Thus, if the prior attainment of the cohort remains similar from one year to the next, then outcomes should also be similar.

Although statistical predictions take priority, there is also a role for judgement in the maintenance of standards (see Chapter 1). Accordingly, the approach aligns with the position of Cizek and Bunch (2007), who suggest that all standard setting processes entail a combination of methods (e.g. statistics and judgement). However, it is the statistical predictions in England that take priority: they drive the process and it is against the statistical predictions that outcomes are reviewed to determine whether standards have been maintained (see Chapter 1). Thus, it can be argued that, based on a strict definition of comparable outcomes (and theoretically what comparable outcomes is trying to achieve), the focus is on the statistical element. As such, it would require convincing judgemental evidence (that is difficult to produce; see Ofqual, 2015b) for grade boundaries to be set that deviate from the statistical evidence. Moreover, when outcomes do deviate from the statistical predictions, it is often due to some technical or statistical evidence that suggests that the statistical predictions are less reliable (see Ofqual, 2017b, c). Given this, throughout this chapter we are considering comparable outcomes in the statistical sense, so we are concerned with the statistical evidence that is used to maintain standards and the implications of EMEE within this context.

Returning to the definitions outlined by Newton (2010a), comparable outcomes therefore implies a causal definition of standards, since the statistical predictions use the causes of attainment – in this case a student's prior attainment – to promote the comparability of standards.[99] This suggests that student attainment in the actual examinations could look different, yet it would still be possible to define comparability in terms of the causes of that

[99] Note that there is some debate about the extent to which standards in England could be defined from a casual or phenomenal definition (see Baird and Gray, 2016). However, given the focus of this chapter on comparable outcomes and the statistical approach to maintaining standards, the adoption of the casual definition seems appropriate here.

attainment (Newton, 2010a). Indeed, one reason for using comparable outcomes is that it protects students in the first year that a new qualification is awarded, when performance is likely to dip – known as the 'sawtooth' effect (Cuff, Meadows and Black, 2017; Linn, 2000; Ofqual, 2015b). Thus, the statistical evidence is prioritised even more when qualifications change (see Ofqual, 2015b).

While the comparable outcomes approach focuses on the prior attainment of students, within the causal definition Newton (2010a) identifies other potential causes of attainment that might be used to infer comparability. These causes are divided into three broad categories, relating to how they might contribute to a student's learning experience. The first set of causes relates to how the students themselves might contribute to their learning experience, the second to how significant others (e.g. teachers, peers or family members) might contribute to students' learning experience, and the third to how learning resources might contribute to students' learning experience. Prior attainment is included in the first of these categories – student causes – yet Newton (2010a) also considers prior attainment to be a special case, since he argues that it is not possible for prior attainment (which is essentially a baseline measure) to have an impact during the course of study in the same way that the other causes might (e.g. teacher support or learning capacity). An alternative interpretation is that prior attainment (and therefore) knowledge can influence future learning: perhaps it is easier to learn if you have certain prior knowledge. However, even if this were the case, prior attainment would still be considered as a student cause so this issue is not discussed further here.

Within the causal definition of standards, Newton (2010a) outlines two sub-categories – the all causes-casual definition, and the specific causes-casual definition. The all causes definition considers comparability in terms of the combined impact of all possible causes of student attainment (see also Cresswell, 1996), while the specific causes definition considers comparability in terms of one or more (but not all) causes of attainment. Given that the comparable outcomes approach prioritises the prior attainment of students and it

is against statistical predictions (based on prior attainment) that grade boundaries are set, this implies a specific causes-casual definition, relating to just one causal factor. Indeed, it is this very fact that has implications for the comparability of examination standards in the presence of EMEE.

Before we consider this further, the following section outlines the changes in entry to GCSE mathematics between 2007 and 2014. Some of this information has been discussed in earlier chapters of this thesis, but it is drawn together here to provide the context for the later discussions about the implications of EMEE for examination standards. First we consider the rise in EMEE between 2007 and 2013, and then the subsequent fall in EMEE from September 2013.

GCSE mathematics entries between 2007 and 2014

The rise in early and multiple examination entry

As outlined previously in this thesis (see Chapters 2 and 4), EMEE to GCSE mathematics rose considerably between 2007 and 2013, meaning that single examination entry at the end of Year 11 was no longer the norm (see Chapter 6). While the qualitative evidence presented in Chapter 5 suggests that this rise was largely driven by accountability pressures and the desire to ensure that students achieved a threshold grade C, there were a number of other contributory factors that facilitated EMEE during this period. These factors primarily related to the availability and structure of GCSE mathematics assessments, and each of these is considered here in turn. First, let us consider the availability of the assessments (see also Chapter 1).

Multiple entries throughout the academic year

GCSE assessments are generally available in a June examination series each year, when the majority of students certificate (or complete) their GCSEs. For GCSE mathematics between 2007 and 2013 there were multiple other examination series available each year,

including a January, March and November examination series. While the availability of each specification from each awarding body tended to change each year, there was at least one GCSE mathematics specification available in each of the four examination series. This meant that in any given academic year students could sit GCSE mathematics at four different points (note that multiple specifications were also generally available in each examination series; see discussion below). This is likely to have encouraged students to enter early and on multiple occasions. In this respect, mathematics was different to the majority of other GCSE subjects where, with the exception of GCSE English, it was generally only possible to certificate once each year (in June).

Multiple specifications from each awarding body

A second factor likely to have contributed to the rise in EMEE between 2007 and 2013 relates to the structure and format of the assessment. GCSE mathematics was unique in structure for a number of years since awarding bodies have typically offered multiple specifications that differ in the format of the assessment. Therefore, between 2007 and 2013, students could choose to sit either a modular specification or a linear specification, with each exam board typically offering qualifications in both formats.[100] The linear qualifications required students to sit all of the assessment in the examination series that they certificated in, while the modular qualifications allowed students to sit (and re-sit) modules prior to the series that they certificated in.[101]

The different formats of assessment increased the number of GCSE mathematics qualifications that were available, effectively doubling the number of specifications compared to the majority of other subjects. This afforded schools and students greater

[100] For the majority of other GCSE subjects either a linear *or* modular format to the assessment was available. Assessment was generally linear at GCSE prior to 2010 and after 2013, and modular between 2010 and 2013.

[101] Note that there was a terminal rule such that students had to sit at least 40% of the assessment in the series that they certificated in and that students could still choose to sit all of the assessment at the end of the course if they desired (see Chapter 1).

choice and flexibility over their entry approaches to GCSE mathematics, and provided more assessment opportunities compared to other subjects. The presence of both linear and modular specifications also made it possible for students to switch specifications partway through the course (see Vidal Rodeiro, 2014). Students could choose to enter the early modules of a modular specification, assess how they were performing, then switch to a linear specification if they were not on course to achieve their target grade (see Chapter 5). This effectively allowed students to 'wipe the slate clean' by nullifying any 'banked' marks from previous modules.

Opportunities to re-sit

A final factor likely to have contributed to the rise in EMEE relates to re-sitting. Between 2007 and 2013 there were no restrictions on how many times students could enter (or re-sit) GCSE mathematics, which series students could enter, and how many specifications students could enter in the same examination series (either with the same or different awarding bodies).[102] Given this flexibility, alongside the availability of the assessments and the desire to ensure that students achieved a particular grade for accountability measures (see Chapter 5), it is not difficult to conceive how EMEE rose in popularity during this period.

The decline in early and multiple entry

Changes to accountability measures

While the rise in EMEE allowed students multiple attempts at achieving a given grade in GCSE mathematics, this approach was not without concern. Several commentators argued that EMEE was not in the best interests of students, but largely stemmed from the pressure on schools to ensure that students achieved a grade C for school performance

[102] The latter was discouraged by scheduling the examinations at the same time on the same day. However, this did not preclude some schools adopting this approach and requiring students to sit the assessments in succession.

measures (see Chapter 2). Thus, in September 2013 it was announced that, henceforth, only a student's first grade would count in school performance tables (DfE, 2013a), a change that would initially apply to the main EBacc subjects,[103] with other subjects to follow from 2016. While students would still receive their highest grade as their final grade, only the first grade would count for the school, requiring schools to balance the needs of the school and the needs of the students in terms of the grades the students achieved (see Chapter 5).

The introduction of this policy had a significant impact on entries in GCSE mathematics. Immediately following the policy announcement a large number of schools withdrew students that they had previously intended to enter (early) to the November 2013 examination series (Ofqual, unpublished; see also Chapter 5), choosing instead to wait and enter their students the following June. Furthermore, the entries to GCSE mathematics in the following summer series (June 2014) from students aged 15 and under (i.e. early entrants) decreased significantly from 170,357 in summer 2013 to just 39,292 in summer 2014 (JCQ, 2014), a trend that continued in subsequent years: in 2015 there were only 33,484 entries to GCSE mathematics from students aged 15 and under, and this fell to 13,209 entries in June 2016 (JCQ, 2015; 2016).[104] Thus, from summer 2014, single examination entry at the end of Year 11 once again became the norm for the majority of students.

The policy change to accountability measures outlined above had a clear purpose to address the rise in early entry, a practice described by Michael Gove, then Secretary of State for Education as *'a damaging trend that is harming the interests of many pupils'*

[103] English, English language, English literature, mathematics, science, history, geography and modern foreign languages.

[104] The sharp decrease in early entry between 2015 and 2016 is also likely to be due to the introduction of reformed GCSE mathematics specifications that will count in performance tables from 2017.

(DfE, 2012). In this respect, the policy could be considered to be successful, since the subsequent changes in entries were firmly in line with the policy intentions. However, as is often the case, the situation is more complex than this. While the implications of the changes in entries for maintaining standards are discussed later in this chapter, there were also consequences for schools and students, leading to criticisms from a number of key stakeholders in the assessment system. For example, the Association of School and College Leaders (ASCL) raised concerns about the damaging effects that 'piecemeal changes' to education were having on students, particularly when changes were made throughout the academic year, as was the case here (ASCL, 2013). They also considered that the policy change would intensify the pressure on schools and students, since students would only have one attempt to achieve a grade for accountability purposes.

The timing for introducing the policy also raised concerns. In a letter to the then Secretary of State for Education, Michael Gove, the Independent Academies Association (2013) appealed for the decision to be delayed so that the current cohort of GCSE students could continue their courses as intended.[105] This view also appeared to be shared within schools, although the evidence for this is largely anecdotal. For example, a number of schools wrote to parents detailing their intentions to change their entry approaches following the policy change, such that the majority of students who had been entered in the November 2013 examination series would be withdrawn. This reflects the discussions reported in the qualitative chapter of this thesis, where the majority of schools reported withdrawing the students that they had entered early for the November 2013 examination series following the announcement (see Chapter 5).

While the change to accountability measures had a significant impact on the prevalence of EMEE, it is important to note that this was not the only change being enacted around this

[105] The policy was introduced in late September, meaning that some students were halfway through their GCSE courses.

time. A number of other changes relating to the availability and structure of GCSE mathematics were also occurring, and to understand the fall in EMEE fully (and the implications for standards, as discussed later in this chapter), requires a consideration of these factors as well.

Changes to the availability and structure of GCSE mathematics assessments

The first change to consider related to the availability of GCSE mathematics assessments. While assessments had typically been available in four examination series prior to 2013, this was reduced following the March 2013 examination series such that there were only two opportunities to certificate each year – June and November. This limited the number of times that students could enter each academic year, effectively reducing the opportunities for EMEE, even if schools still wanted to adopt this approach following the changes to accountability measures.

A second change related to the 'rules' for entering the November examination series. From 2014, the November examination series became a re-sit only opportunity, meaning that students were only eligible to enter if they had also entered the previous summer. Prior to this, students could enter for the first time in November, and many chose to do so.

A third change concerned the structure of GCSE assessments. From summer 2014 all GCSEs had to be sat in a linear format (i.e. all assessments were taken in the series that students certificated in), while prior to this many GCSEs had been modular, meaning that students could sit part (but not all) of the assessment prior to the series in which they certificated. The move to linear assessments was first signalled in the Importance of Teaching White Paper (DfE, 2010), following concerns that modular assessments led to over-assessment and re-sitting that undermined the qualifications. Thus, although the structure and content of GCSE qualifications remained the same, all assessments had to be taken in the same series. For GCSE mathematics, where linear specifications were already available, students could therefore choose to sit the existing linear specification or sit the (formerly) modular specification linearly.

In summary, the changes to the availability and structure of GCSE mathematics assessments, coupled with the changes to accountability measures, had a significant impact on the way that students entered GCSE mathematics assessments. In turn, this had implications for examination standards. From an operational perspective, the shifts in entries created challenges for maintaining standards, both as the prevalence of EMEE increased and declined. The following section discusses the implications of these two shifts in entries for maintaining standards, given the causal definition of standards that has been outlined above. First, let us consider the implications as EMEE increased.

Implications for examination standards

The increase in EMEE between 2007 and 2013 meant that rather than entering GCSE assessments once, many students were entering on multiple occasions and therefore certificating multiple times and achieving multiple grades. The following discussion considers the implications of this for the comparability of examination standards, given how standards have been defined (see earlier discussion in this chapter).

How can standards be maintained when early and multiple entry increase?

Earlier in this chapter we discussed that the comparable outcomes approach implies a specific causes-casual definition of standards: one would consider standards to be comparable if the inputs to learning – in this case students' prior attainment – are comparable (Newton, 2010a). Thus, there is an expectation that students at linked grade boundaries will be comparable in terms of their prior attainment, i.e. a student achieving a grade C across two different years will have the same cause of their attainment. This definition provides the criterion against which we can consider the implications of EMEE.

It has already been discussed in Chapters 6 and 7 that students entering on multiple occasions will likely achieve different grades on different occasions, sometimes due to the unreliability that is present in assessment, and sometimes due to other factors such as

maturation or teaching. Crucially though, a student will always receive their highest grade as their final grade. In this way, it has been argued that EMEE will likely increase outcomes, since students can only ever achieve a higher grade through EMEE (see Chapter 6). While some of the increases in performance from one series to the next will be due to legitimate improvements, others will likely be due to the error that is inherent in assessment. The latter may result in a student achieving a grade that is not a reflection of their 'true' score (see Chapter 6).

To illustrate this and the implications of EMEE for standards, consider the following example of two students with the same prior attainment who follow different entry routes – one student enters once, and the other enters on multiple occasions. Compared to the former student, it is much more likely (though not a given) that the latter student will achieve a higher grade on the assessment, purely because they have entered multiple times (see Chapters 6 and 7 and Millman, 1989; Wheadon, 2010). This is illustrated in Figure 21. This raises questions around comparability given how we have defined examination standards: although comparable outcomes aims to ensure that once prior attainment has been taken into account outcomes are similar, in the presence of EMEE, students with the same final grade will likely have different causes of their attainment – the single entry student will likely have higher prior attainment than the multiple entry student. Thus, although comparable outcomes aims to achieve comparability based on prior attainment through the use of statistical predictions, EMEE and the cumulative effect of this over time compromises this, and raises questions around the extent to which standards can be considered to be comparable. Consequently, while one might consider standards to be comparable when considering just a single examination series, across

multiple examination series when students have entered multiple times, this definition is unlikely to hold.[106]

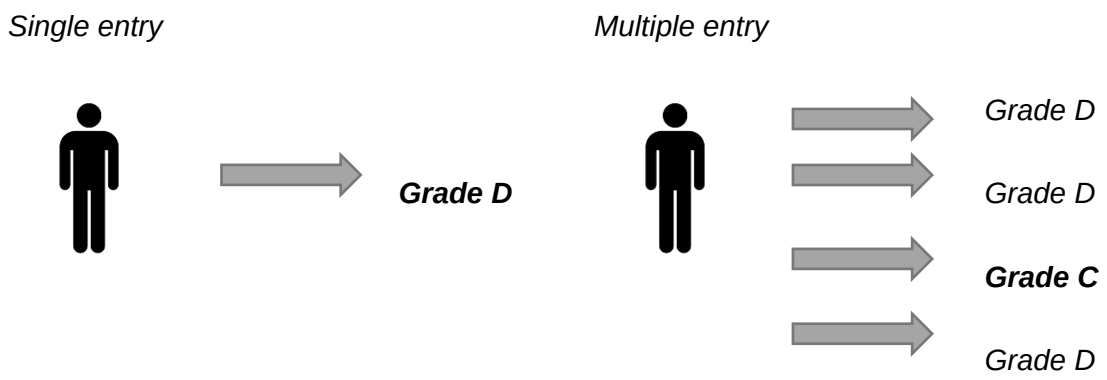


Figure 21. *Outcomes for single and multiple entry students with the same prior attainment*

In practice, the situation is much more complex than this. While the discussions above have considered the distinction between students entering once and those entering on multiple occasions, in practice there are likely to be several routes through multiple entry. As shown in Chapter 6, some students entered twice before the end of Year 11, yet others entered on up to 12 occasions. This is likely to have different implications for the grades that students achieve, since the greater the number of entries, then the greater the likelihood that students will achieve a higher grade (see Chapter 7 and Millman, 1989; Wheadon, 2010).

Given that there was a sharp rise in EMEE for GCSE mathematics between 2007 and 2013, what, therefore, are the implications for standards? From the discussion above it is becoming apparent that comparable outcomes might not be sufficient to ensure comparability when both single and multiple examination entry are simultaneously in use

[106] There are also likely to be implications over time for successive cohorts of students, but these are not discussed in detail here. For example, if multiple entry increases outcomes due to the error inherent in assessment, this will be built into the predictions for subsequent cohorts. There might therefore be a cumulative effect over time.

(or indeed when only EMEE is in use), assuming that standards are considered from a specific causes-casual perspective (i.e. the definition that comparable outcomes implies; see Newton, 2010a). Comparable outcomes uses only one cause of attainment – prior attainment – yet it seems that this alone cannot ensure comparability in the presence of EMEE. Assuming that outcomes meet the statistical predictions each examination series then comparable outcomes might seem sufficient at face value, but over time it is not. This raises the question of whether other causal factors need to be taken into account to ensure that examination standards are maintained – for example, is it the case that further information about each student is needed, as might be suggested by the catch-all definition of standards outlined by Cresswell (1996)? Or if not, can we only think about standards as comparable for those students who followed the same entry route through the assessments – for example, a student that entered once could only be compared with another student that entered once?

In his work, Newton (2010a) described a number of causal factors that might be used to define comparability (see earlier discussion in this chapter), but does the case of EMEE imply that the 'entry' approach of students is another causal factor? In some ways, it is difficult to conceive how entry approach could be considered as a casual factor, since it cannot influence learning during the course of study. Furthermore, unlike prior attainment, it is not a baseline measure. However, one could perhaps argue that is it a causal factor in the sense that it influences a student's final attainment, and can therefore have implications for comparing students at linked grade boundaries. Thus, while the comparable outcomes approach might be sufficient when entries to assessments are stable, in the presence of EMEE comparable outcomes seems insufficient to ensure comparability in terms of the causes of students' attainment at key grade boundaries.

It is therefore worth considering what would be required to ensure that, within the context of EMEE, standards are maintained. For example, if it is assumed that every student will enter an assessment twice (rather than once), then should the approach for maintaining

standards reflect this to ensure that multiple entry does not comprise comparability? Achieving this would not be straightforward and there are operational considerations of trying to do this (whether logistically it is possible) and policy implications (whether it would be appropriate to do this). These are considered below as we discuss the possible ways in which one could account for the presence of EMEE in GCSE assessments, to ensure that it does not comprise the comparability of standards assuming Newton's (2010a) casual definition. The discussion focuses on the approach that is used to maintain standards in England (i.e. statistical predictions) because the statistical evidence drives the process through comparable outcomes (see Chapter 1). However, since there is also a role for examiner judgement of student performance (see Chapter 1), the implications of these approaches for judgement are also considered.

The following discussions have been situated within the context of the assessment system in England at the time that EMEE was widespread. While it is acknowledged that there could be other solutions for ensuring comparability in the presence of EMEE based on approaches from other jurisdictions (e.g. an item response theory [IRT] approach perhaps), adopting such an approach in England would pose both technical and ethical challenges (see Wheadon, 2011), and it is not immediately obvious whether such an approach would solve the issue. Furthermore, while any technical and ethical challenges are unlikely to be insurmountable, any changes would be unlikely to occur quickly because of this (Wheadon, 2011), and certainly not within the timeframe that the prevalence of EMEE rose. Thus, possible alternative approaches to maintaining standards are not discussed further here and are considered to be beyond the scope of this thesis.

Approaches for maintaining standards in early and multiple examination entry

The final grade counts

One possible method of mitigating the effects of EMEE would be to only allow students to retain a particular grade as their final grade – e.g. to state that students should receive the last grade that they achieve as their final grade, meaning that any subsequent grades

would replace a previous grade. Thus, a student who entered on multiple occasions and legitimately improved their grade through maturation or further teaching would benefit. However, students that achieved a different grade because of the unreliability of assessment would not benefit (assuming that the higher grade was not achieved on the last occasion, though this would likely happen in some cases).

While this policy would be easy to operationalise (but rely on the exam boards to do so), this would no doubt be a highly controversial and unpopular policy, since it would likely penalise some students who, for whatever reason, performed poorly on their final attempt. It would also be difficult (if not impossible) to justify, since students might have performed at a higher level in the past than their final grade would suggest. It would, however, reduce the potential for students to increase their grade through entering on multiple occasions – it is the fact that students receive their highest grade as their final grade that can result in outcomes increasing across multiple entry attempts (see Chapter 6). Indeed, it may serve to discourage EMEE by its very nature since the stakes associated with the final entry attempt would be much higher, and the incentive to enter multiple times (i.e. that the best grade counts as the final grade) is reduced. Alternatively though, it may encourage EMEE further since students might continue entering until the desired grade was achieved, raising further concerns for comparability in Newton's (2010a) causal sense.

Limiting or not allowing re-sits

Another approach might be to limit (or remove) the opportunity to re-sit a qualification.[107] Wheadon (2010) showed that as the number of re-sits increased, the chance of students

[107] Note that in higher education in the UK the re-sits of individual units are sometimes capped at a pass (generally 40%; Ricketts, 2010). Given that passing every assessment in a GCSE qualification is not a requirement a similar approach would not be possible so it is not discussed further here. Furthermore, GCSEs have adopted a linear structure since 2014 so again this approach would not be feasible (see Chapter 1).

being misclassified and achieving a higher grade than their 'true' score might suggest also increased (see also Millman, 1989). This suggests that limiting the number of re-sits would be one method to reduce the potential for multiple entry to increase students' grades, although consideration would need to be given to the number of re-sit opportunities that would be allowed (and a strong rationale for any decision required): the greater the number of re-sit opportunities then the greater potential effect on outcomes, the fewer the number of re-sit opportunities then the smaller potential effect. Furthermore, while limiting the number of re-sit opportunities might reduce the number of false positives, it would do nothing to protect a student from a false negative – that is, achieving a lower grade than their 'true' score would suggest (see Clauser and Wainer, 2016; Millman, 1989).

Having very few or indeed no re-sit opportunities would also be controversial and difficult to justify, and would raises issues around fairness: it might penalise students who performed poorly on their first attempt (for whatever reason), including those students who were mis-classified below their true grade initially due to random assessment error (see Millman, 1989; Wheadon, 2010). It might also increase the pressure on students to perform. In practice, only having one entry attempt would also pose a significant barrier to accessing further education and employment opportunities for those students that did not achieve a grade C on their first attempt, and would be at odds with current Government policy that requires full-time students to be working towards achieving a good pass in English and mathematics (if they did not achieve this aged 16) during post-16 education for funding purposes (see Education and Skills Funding Agency, 2017). However, from an operational perspective it would be relatively straightforward to implement (again relying on the exam boards) and might, in itself, discourage early entry since there would be no (or limited) opportunity to re-sit – entering students before they were ready would be much more undesirable.

Manipulating the statistical evidence

Alternative methods to mitigate the effects of EMEE on the comparability of standards would be to account for the entry approaches used by students through the approach used to maintain standards. In the case of GCSEs, this would entail manipulating the statistical evidence that is used to support the setting of grade boundaries – the statistical predictions. Schools would then retain choice over when to enter their students, but this would be accounted for in the grades that were awarded. For example, if one could demonstrate that multiple entry increased outcomes by a given percentage at each grade across examination series (based on previous data), then the statistical predictions could be adjusted to account for this. This would aim to ensure that standards were maintained in terms of the causes of student's attainment (see Newton, 2010a), when all of a student's results were taken into account. Thus, the causes of attainment would extend beyond a student's prior attainment. This would not satisfy the catch-all definition proposed by Cresswell (1996), since we would not take account of all the factors that might influence attainment, yet comparability could still be considered in terms of a specific causes definition of standards as outlined by Newton (2010a; see earlier discussion in this chapter). Now though, there would be multiple causes of attainment.

This may seem like an appealing solution since it would promote comparability in terms of the causes of attainment in the presence of EMEE, yet operationalising this would be complex. This is particularly the case when it is considered that some increase in outcomes due to EMEE would be due to legitimate improvements in performance (e.g. due to student preparedness or motivation), meaning that quantifying the extent to which multiple entry (in itself) influenced outcomes (ignoring any real improvements) would be challenging. For example, it would first require one to identify the factors that yield real improvement in performance and quantify their effects. This would be challenging since it is not merely a statistical problem that might be solved by comparing the performance of students on different entry attempts. Rather, one would need some sort of method for

measuring the impact of factors that legitimately improve performance such as test preparedness, quality of teaching and student motivation, so that any effects of measurement error were isolated. Given that there are differences in how schools prepare their students for assessments in the presence of EMEE (both within and between individual schools; see Chapter 5), it is difficult to conceive how any measure of preparedness is possible. Furthermore, while measuring something like student motivation would be more straightforward (and there are measures to do so that have been applied in the GCSE context; Putwain and Remedios, 2014; Vallerand, Pelletier, Blais, Brière, Senécal and Vallières, 1992), a measure of motivation would presumably rely on some sort of self-report data. This might not be sufficiently reliable, particularly when it is considered that any adjustments to the predictions would need to be fairly precise and publically defensible.

Even if we were to assume that it was possible to quantify the effects of factors that lead to real improvements in performance and remove those, quantifying the effects of EMEE and implementing any adjustments would still be challenging. This might be relatively straightforward if an assessment was only available once each academic year and all students sat the same qualification, but this is not the case in England (see Chapter 1). At the point when EMEE was at a peak, GCSE mathematics assessments were available up to four times each academic year, each time with multiple awarding bodies (see Chapter 1). Furthermore, within a given awarding body multiple mathematics qualifications were available (e.g. a linear and modular GCSE, a linked-pair pilot qualification comprising two GCSEs, and an international GCSE or level 1/2 certificate; see Chapters 1 and 2). As such, quantifying any adjustment and indeed making any sort of adjustment to the statistical predictions (across multiple series, providers and qualifications) would be significantly more complex.

It is also feasible that due to differences in the entry profile within each exam board (e.g. if one board had no multiple entrants and another had many), different adjustments would

be needed. This would introduce additional complexities into the process. It would also be difficult to explain to stakeholders, particularly if the adjustments differed across exam boards and grades (any differences would likely be construed as one exam board being made ‘easier’ than another). Indeed, explaining the rationale for the approach itself would be challenging, since this would require an understanding of the concept of assessment error, which is not widely understood by stakeholders (see Chamberlain, 2013 and Chapter 6).

A further complication in calculating any adjustments is that there are not just single and multiple entry students: students that enter on multiple occasions do some in many different ways (see Chapters 5 and 6), and this would also need to be taken into account. This would complicate any calculations considerably, if they were indeed feasible at all.

A final issue is that this approach might advantage or disadvantage the ‘wrong’ students (see also Millman, 1989). As predictions are calculated at the overall cohort level (see Chapter 1) any downward adjustments might have a knock-on effect for those students that were not entering on multiple occasions, who might consequently achieve a lower grade. This would be unfair, but difficult to mitigate, unless single and multiple entry students were treated differently.

Distinguishing between first-timers and multiple entrants

This leads us to a similar but alternative suggestion: to separate students into groups based on the number of times that they had entered a qualification in a particular subject, then treat each group differently in the process used to maintain standards. This would aim to ensure that overall comparability was achieved in terms of the causes of students’ attainment (i.e. their prior attainment; see Newton, 2010a). For the students that had only entered once the comparable outcomes approach should be sufficient, while for the others some sort of adjustment would be needed. In other words, students who were entering on multiple occasions would have to demonstrate a higher level of performance to achieve a

given grade on subsequent entry attempts, to account for the fact that entering on multiple occasions would likely increase outcomes due to assessment unreliability.

This approach has been discussed elsewhere in relation to licencing examinations. Millman (1989) suggested that pass marks on re-sit attempts could be increased to ensure that test-takers do not pass an assessment purely through having entered on multiple occasions. Given that the grade boundaries for GCSEs change each series (see Chapter 1), then increasing the pass mark directly would not be feasible in quite the same manner, yet adjusting the statistical predictions would achieve this by having a knock-on effect for the grade boundaries.

The practice of adjusting pass marks has also been discussed in relation to medical assessments (Ricketts, 2010; see also Clauser and Wainer, 2016), where the stakes are arguably higher: while GCSE results can allow students access to further education and employment opportunities (see Chapter 1), the implications of individuals who are not fully competent passing licence to practice medical assessments are likely to be far more serious. However, even in this arena it is acknowledged that having sufficient information to know how much a pass mark should be moved by is not straightforward (Clauser and Wainer, 2016; Ricketts, 2010), and that the very act of increasing the pass marks is likely to be controversial (Pell, Boursicot, and Roberts, 2009). These are similar challenges to those that we would face here in relation to EMEE.

While our suggested approach of differentiating between students when maintaining standards would protect the single entry students from being unduly penalised, operationalising this would again be problematic. It would require students to be reliably categorised based on the number of entry attempts that they had made, meaning that co-operation and data sharing between exam boards would be needed. This is not insurmountable, but would introduce some complexity into the process. Like the previous approach there would also be concerns around explaining it to stakeholders, and questions around fairness (see also Millman, 1989). Moreover, while each student would

be grouped with other students according to how many times they had entered a qualification, within each group the individual students are likely to be different. While some students that are entering on multiple occasions might be doing so due to the pressures on the school (see Chapter 5), others might have performed poorly on an earlier attempt due to illness or personal circumstances. It would seem unfair to treat these students the same, but distinguishing between them reliably would be challenging.

Another issue relates to how any adjustments would be operationalised. Using different predictions for each group and setting different grade boundaries would be one option, but this would be complicated and difficult to explain to stakeholders. Indeed, generating predictions for the multiple entry students would entail the same challenges as outlined previously. Adjusting students' marks (while maintaining the same grade boundaries) would be even more difficult, since the number of marks to achieve a given grade differs each examination series. Thus, basing any adjustments on data from a previous examination series would not be appropriate, making it is difficult to conceive how any mark adjustments could be calculated. Justifying this approach would also be difficult, since adjusting students' marks would change the rank order of students. This would mean that students who had achieved a higher mark but had entered multiple times would achieve a lower grade in a particular examination series than students who had performed worse but had entered once. The extent to which this would be publically defensible would be questionable.

Reporting grades differently

A more straightforward approach might be to not attempt to adjust the statistical predictions and allow students to retain their highest grade as their best grade, but to report students' results according to whether they were re-sitting or not. This would be fairly straightforward to operationalise and would allow schools choice over their entry policies. However, without any guidance about how this information should be interpreted it is unlikely to be useful to users of the qualification results, and any such guidance would

be difficult to generate since each individual case is likely to be different. This would mean that the users of the qualification results would be left to decide whether they wished to take account of how many entries students had made when interpreting results. Some stakeholders may decide to prioritise students that had made fewer entry attempts, though this could penalise some students unnecessarily and might therefore raise issues around fairness. Others would likely not take this information into account, leaving students uncertain how their results would be interpreted. Again, this seems problematic.

In addition, and perhaps more importantly, this would not solve our original issue: while it might aid the interpretation of results, results are still unlikely to be comparable in terms of the causes of students' attainment – in this case, their prior attainment (see Newton, 2010a). This is because a student would still achieve their final grade as their best grade, and this would be used in subsequent years to generate predictions for later cohorts of students. Thus, while reporting results differently would allow users of the qualification results to interpret them differently, it would not assure us that standards had been maintained given our causal definition (see Newton, 2010a).

The implications for examiner judgement

The discussions above suggest that accounting for EMEE statistically or through policy in the context of the assessment system in England is not straightforward, yet it is also worth considering the role of examiner judgement for maintaining standards (see Chapter 1) and the implications of the suggestions outlined above for this. For some of the suggested approaches there would be few implications for examiner judgement, as the approach to setting grade boundaries in a given examination series would not change – for example, if policy were to dictate the grade that counted was a student's final grade, or the number of re-sit opportunities were capped. However, if adjustments were made to the statistical evidence then there would be implications, and this is because of a point discussed earlier in this thesis (see Chapter 6). That is, when considering any one given examination series, standards would appear to be comparable even in the presence of EMEE (if

defined in the causal sense; see earlier discussion in this chapter and Newton, 2010a), since a student only has one grade (and any unreliability inherent in assessment would cancel out across the whole cohort). However, across multiple series this is not the case (see Chapter 6). As such, if adjustments were made to the statistical predictions in one series (due to EMEE), examiners would be required to account for this when making qualitative judgements of student's work, despite in that particular examination series there being no apparent effect of EMEE – i.e. they would be expected to account for what might be described as the 'bigger picture'. Given the evidence showing the limitations of examiner judgement (see Chapter 1) this would not be easy, suggesting that the statistical evidence would need to be prioritised, perhaps even more so than it is currently (see Chapter 1 and Baird and Gray, 2016). However, as described earlier in this section, this is not straightforward either.

Can multiple entry help us find a students' most accurate grade?

The discussion above has suggested possible approaches to ensure that, in the presence of EMEE, comparability of standards is maintained in the causal sense (see Newton, 2010a). This suggests that EMEE is a bad thing that comprises comparability. But what if we consider EMEE differently – as a mechanism for allowing us to derive the most accurate grade for each student? The nature of EMEE means that students are sitting more assessments than if they just entered once. They are therefore being tested on more of the syllabus (though there would likely be some overlap) and, in this way, EMEE could be considered as a longer version of a single GCSE. If all of the assessments were taken into account when arriving at a final grade, through some sort of averaging, then in general this should increase the reliability of the assessment. It would also promote comparability in terms of the causes of attainment since students would not be achieving higher grades through assessment error – this would be averaged out (assuming that sufficient entry attempts were made by each student).

While theoretically this might sound appealing, operationalising this approach would be challenging. Aggregating all of the assessments that each student took together by averaging the marks, for example, would not be possible, since individual students would have a unique combination of assessments taken at different time points when the demand of the assessments differed. While it would be possible to use a conversion such as UMS (see Chapter 1) to place the marks on a standardised scale, this would require a set of grade boundaries that would need to be determined for each assessment. This means that grades would still need to be awarded separately each examination series based on the performance of those specific assessments, so averaging marks would not help us overall.

The alternative approach would be to average the grades that students achieved across all of their entry attempts and report the average as the final grade (e.g. see Millman, 1989). This draws some parallels with an approach that has been discussed in the higher education context, where a student's prior performance would be taken into account to aid the interpretation of performance on the re-sit (see Ricketts, 2010). What we are suggesting here though is that the grades would be awarded separately then averaged.

Since GCSE grades are awarded as letters these would need to be converted to a number, then the average taken. This would be straightforward to do, but the interpretation would be difficult. Students' grades could either be reported on a decimal number scale[108] or their average grade could be rounded. The former might suggest a greater level of precision in the assessments than is realistic, which might encourage users of qualification results to make decisions based on very small differences between

[108] This would negate the use of letter grades, since it would only be possible to convert whole numbers back to letter grades. This is not necessarily problematic though since GCSE assessments are moving towards a 9 to 1 grading scale from 2017 anyway.

students.[109] The latter, however, might ensure that if a student achieved a higher grade as a 'one-off' (for example, through measurement error; see Chapter 6), then this would have limited influence on their final grade. For example, if a student achieved four grade D's and one grade C, their final (rounded) grade would still be a grade D.

If a method of averaging grade across multiple attempts was used there is another challenge relating to the purpose of results though, particularly in relation to how an average grade would be interpreted in the context of students re-sitting to meet a threshold measure, such as a grade C. Consider a student that achieved a grade D on their first attempt then a grade C when they re-sat in post-16 education, through legitimate improvements in their performance. If averaging (with no rounding) was in effect, their final grade would be 4.5 (assuming that A* – G grades were converted to 8 – 1 grades). Technically, they would not have achieved the required threshold if this average took precedence, though they would have demonstrated the required level of performance once. This means that it would be difficult to compensate for poor early performance for those students that genuinely improved at a later date, suggesting that this approach would be inherently unfair if universally applied. However, trying to treat students differently would be challenging since there would be nuances in every situation.

A final point to consider is that while theoretically the more entry attempts that each student made the better (for reliability and if grades were averaged), in practice this approach would likely discourage EMEE since entering students before they were ready (when they would be more likely to under-perform) would be undesirable. This might prevent the approach achieving the desired gains – a more reliable assessment result. Discouraging EMEE might not be a bad thing but because of this, it is unlikely that EMEE would ever allow us to achieve a more reliable assessment result.

---

[109] The possible values that each student could achieve would also not be comparable since it would depend on the number of entry attempts that they made.

Summary of possible approaches

In summary, this discussion has outlined some of the complexities in trying to account for changes in entry trends – particularly the rise in EMEE – when maintaining standards via the comparable outcomes approach and defining standards in the casual sense (see Newton, 2010a). This is further complicated when it is considered that individual students are likely to take their own route through qualifications. As shown in Chapter 6, students entered GCSE mathematics anywhere between one and 12 occasions before the end of the 2012 – 2013 academic year, and for those students entering twice, there were 40 different routes through the qualifications (without even accounting for which specification or awarding bodies students entered with). It is clear that ensuring comparability in the presence of EMEE is not straightforward.

In the case of EMEE, these discussions are somewhat theoretical since EMEE in GCSE mathematics declined following policy changes stipulating that only a student's first grade would count for school performance measures from September 2013 (yet students' could re-sit to improve their grade; see DfE, 2013a). This policy change effectively ceased EMEE due to the potential implications for schools' performance measures (see Chapter 5), and while the approach is similar to the suggestion outlined above that only the final grade that a student achieves should count, it is more nuanced. For the individual student the highest grade still counts, yet for the school only the first grade counts. This effectively discourages early entry before students are ready since there will be implications for school performance measures, yet it does not penalise the students where EMEE does occur (for whatever reason). While the reduction in EMEE reduced the potential for multiple entry to impact on outcomes and the comparability of standards, the shift back from multiple entry to single entry between academic years had its own implications for examination standards – and it is to this issue that the chapter now turns.

How can standards be maintained when early and multiple entry declines?

The second issue for maintaining standards relates to the decline in EMEE in 2013. As outlined previously, EMEE to GCSE mathematics was a popular approach prior to 2013, yet this changed from November 2013 onwards following changes to accountability measures and changes to the availability and structure of GCSE assessments (see earlier discussion in this chapter). This meant that the cohort sitting GCSE mathematics changed considerably between summer 2013 (when students were entering early and on multiple occasions), and summer 2014 (when single examination entry became the norm). This change created challenges for ensuring comparability of standards in summer 2014, primarily because the type of students certificating in summer 2014 were not comparable to the students certificating in summer 2013, yet the statistical predictions upon which comparable outcomes is based relies on the assumption that the students – and therefore their value added – are similar in the reference series and the current series (see Chapter 1).

The decline in EMEE was a live issue and the discussion below outlines the considerations undertaken by the exam boards and Ofqual at the time EMEE declined, and the decisions that were taken. However, to extend the discussion beyond simply recounting the approach taken in practice, the discussion is situated within the theoretical literature relating to how examination standards are defined (see earlier discussion in this chapter). Before considering the implications of the decline in EMEE for maintaining standards, let us consider the changes in the cohort taking GCSE mathematics between 2013 and 2014 in more detail and the implications of this.

Changes to the cohort

The first difference in the population of students sitting GCSE mathematics in summer 2013 and summer 2014 relates to the age of the students. In summer 2014, the proportion of students entering early (aged 15 and under) decreased considerably compared to the previous year, from 170,357 in summer 2013 to just 39,292 in summer 2014 (JCQ, 2014).

This followed a sharp decline in early entries in the November 2013 examination series (see also Chapter 5). The decrease in entries from 15-year-old students in summer 2014 meant that the age profile of the cohorts was very different between summer 2013 and 2014, as illustrated in Table 25.

Table 25. *Age profile (%) for GCSE mathematics cohort in summer 2013 and 2014 (JCQ, 2014)*

	Aged 15 and under	Aged 16	Aged 17 and over
Summer 2013	22.41	67.39	10.20
Summer 2014	5.34	81.01	13.66

A second difference relates to the number of multiple entry students, which would have decreased in 2014 relative to the previous year. To have entered multiple times prior to summer 2014 students would have needed to have certificated in a previous series or be entering two specifications in the same examination series. Given the decline in early entry in November 2013 and the reduced availability of assessments from March 2013, there were fewer opportunities to enter early for those students who ultimately certificated in summer 2014. As such, it is less likely that students certificating in summer 2014 would be re-sitting from a previous series, compared to the students certificating in summer 2013. This meant that the majority of students entering in summer 2014 were likely to be entering for the first time.

The reduced availability of assessments also had another effect. Students that had previously entered early and chosen not to re-sit in the summer series would now be part of the summer 2014 cohort.[110] This might have been students who had achieved or exceeded their target grade in an earlier examination series, or students who were not

[110] Note that some students may still have not entered in a summer series, but the numbers would likely be considerably smaller.

motivated (or permitted) to re-enter in a later examination series to improve their grade (see Chapters 2 and 5).

A third change related to the structure of GCSE mathematics assessment. Following changes such that all GCSE assessments had to be sat in a linear format from summer 2014 (see Chapter 1), the GCSE mathematics cohort shifted from the (formerly) modular specification to the linear specification. This is shown in Table 26, which shows the percentage of entries to each specification for the two exam boards with the largest market shares in GCSE mathematics in summer 2013 and 2014.[111] In many ways, the increase in entries for the linear specifications is not surprising since this assessment had always been designed to be sat in a single series, while the modular assessment had been designed to be sat across multiple series. As such, there were a greater number of assessments in the modular specification (three) compared to the linear specification (two).

Table 26. *Entries (%) to modular and linear specifications summer 2013 and 2014*

	Edexcel		AQA	
	Modular	Linear	Modular	Linear
Summer 2013	12.37	87.63	43.45	56.55
Summer 2014	1.13	98.87	11.81	88.19

These changes show how the cohort of students entering in summer 2014 were likely to be very different to those entering the previous summer, the series that generally would be used as the reference year for generating the statistical predictions (see Chapter 1). The differences apply to the entire cohort, but also for specific age groups. The cohort of students aged 15 and under would likely have been very different between 2013 (when

[111] Note that this just refers to the two main GCSE mathematics qualifications where the majority of entries were made, not the linked pair pilot or any level 1/2 certificate qualifications (see Chapters 2 and 6).

entries were large) and 2014 (when entries were small), since schools were only likely to enter certain students early in 2014 – probably the more able. More importantly, the cohort of students aged 16 (upon whom awarding decisions and grade boundaries are set; see Chapter 1) would also be very different – fewer would have entered in a previous series and be re-sitting, and students who previously would have certificated early then not entered in the summer series would do so in summer 2014.

Generating statistical predictions in summer 2014

The shifts in entries, influenced by policy changes and changes to the availability and structure of the assessments, raised questions for the process of maintaining standards in summer 2014 via the comparable outcomes approach, assuming a causal definition of standards (Newton, 2010a). If the 16-year-old cohort in summer 2014 were not comparable to the 16-year-old cohort in summer 2013, then how should statistical predictions be generated given that some sort of statistical evidence is necessary to facilitate the maintenance of standards (see Chapter 1 for a discussion)? Furthermore, how can comparability, as defined as the causes of students' attainment – their prior attainment (see Newton, 2010a) – be assured?

Following the comparable outcomes approach would suggest that outcomes in the summer 2014 series should be predicted for 16-year-old students, based on the outcomes of 16-year-old students in summer 2013 (or an alternative preceding June series). However, as outlined previously, the cohorts, and indeed the 16-year-old students, were very different – both between summer 2013 and 2014, and between summer 2014 and earlier summer series – meaning that the value added of the students each year was unlikely to be comparable. This posed a risk to the comparability of standards (see Newton, 2010a).

Approaches to maintaining standards

One option for maintaining standards would be to use a different age group of students to generate statistical predictions in June 2014 – i.e. one that was more stable between 2013

and 2014 than the 16-year-olds. However, this was not straightforward either. Basing predictions on 15-year-olds would be unfeasible since the cohort of 15-year-olds changed considerably between 2013 and 2014, perhaps even more so than the 16-year-olds. Thus, the students that entered early in 2013 aged 15 – when there were many – were unlikely to be comparable to the students entering early in 2014 aged 15 – when where were relatively few (see JCQ, 2014). Furthermore, the 15-year-old students would be a minority age group, suggesting that any statistical predictions might be less robust if the numbers were small (see Chapter 1).

An alternative option was to base the predictions on older students – students that are in post-16 education. However, post-16 students are not necessarily a stable group either – both in the context of changes in EMEE, or when entries are more stable: they contain a mix of age groups, some of whom are entering for the first time, and some of whom are re-sitting. Thus, these different groups of students are likely to perform very differently. This, coupled with the issues of using 15-year-olds, suggests that generating predictions based on alternative age groups in summer 2014 was not feasible. Indeed, generating predictions for any age group other than 16-year-olds in a summer examination series is likely to be difficult to justify and defend, since they would include lower numbers of students that would likely result in less robust statistical evidence.

Given that generating predictions for a different age cohort was not feasible, an alternative approach might have been to generate predictions based on an alternative reference series, when the students were more comparable to those certificating in summer 2014. However, a brief consideration of the possibilities suggests that this was unlikely to be feasible either. While there were several examination series in 2013 – January, March, June, and November – none of these were likely to offer an acceptable comparison with summer 2014. In the January and March series only certain specifications were available, which might have influenced the type of students that entered. Moreover, the January, March and November are all at different time points to the summer series, meaning that

students might have performed differently based on their levels of preparedness for the assessment. Again, this would not assure comparability in terms of the causes of students' attainment.

A final alternative was to dispense of the standard approach for generating statistical predictions based on the same series in the reference year (see Chapter 1), and to generate predictions for an entire academic year instead. This would mean that all types of students entering in any examination series in a previous academic year – in this case 2012 – 2013, since this was the most recent – would be included in the base of the predictions. This has the advantage of avoiding prioritising one examination series over another as the reference series, and is likely to provide greater parity in the type of students between the reference year and the current year. Furthermore, it has the advantage of increasing the number of students included in the predictions, since they would include all entries across each academic year.

In practice, this approach was adopted by the awarding bodies since it was considered most likely to mitigate the risk that changes in the cohorts would compromise the comparability of standards in summer 2014. Thus, statistical predictions based on outcomes for all students in any examination series across the entire previous academic year (2012 – 2013) were used to predict outcomes across the whole of the current academic year (2013 – 2014; see Ofqual, 2014b).[112] Therefore, if the prior attainment profile of the students across each academic year was similar, then the predictions would aim to ensure that the outcomes across each academic year were also similar. This is essentially what comparable outcomes aims to achieve and therefore aligns with the causal definition of comparability outlined by Newton (2010a; see previous discussion in

[112] Note that as the November 2013 series had already been awarded the outcomes from November 2013 were taken into account when generating statistical predictions for the summer 2014 series, i.e. the whole academic year predictions were a combination of November 2013 and summer 2014.

this chapter), where comparability is defined in terms of the causes of attainment – in this case, prior attainment.

Multiple entrants – first, best or last grade?

While the approach of using whole academic year predictions was intended to support the maintenance of standards and ensure that students who certificated in 2014 achieved comparable outcomes to those certificating in 2013 (once their prior attainment had been taken into account), it did create a second challenge. Given that the predictions were to be based on the whole academic year and that many students in 2012 – 2013 had entered multiple times, there was a question about which result should count for multiple entry students in the basis of the predictions (see Chapter 1 for a discussion of how predictions are generated). Although students receive their highest grade as their final grade, using this in the basis of the predictions would mean that any increases in outcomes that were due to multiple entry and the inherent unreliability in assessment (see Chapters 6 and 7) would be carried forward to the 2014 cohort via the statistical predictions (see earlier discussion in this chapter). For example, a student that had entered four times in the 2012 – 2013 academic year and achieved grades C, C, B, C, would have the grade B used in the basis of the predictions, as illustrated in Figure 22.

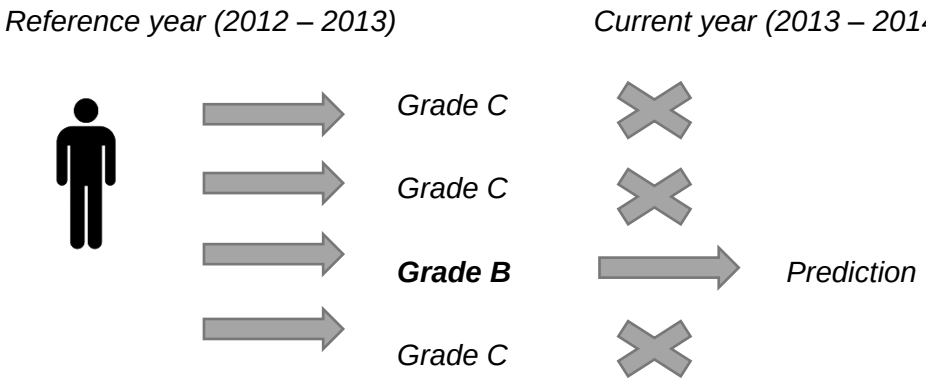


Figure 22. *Prediction based on a student's final (i.e. highest) grade*

Given the concerns around EMEE and the policy changes to discourage this behaviour (see earlier discussion in this chapter), there is a question of whether this would be appropriate. This issue is somewhat unique to this situation, since typically predictions are based on only one reference series, where the vast majority of students will have only entered once. Therefore, there is no precedent to follow. Let us consider the various options available.

To generate statistical predictions requires the relationship between prior attainment and outcomes to be modelled in a reference series (see Chapter 1). While it had already been decided that in summer 2014 the references series should be the whole 2012 – 2013 academic year (accounting for the November 2013 results that had already been awarded), it was noted that for any students that entered on multiple occasions, the relationship between prior attainment (mean KS2) and their outcomes (GCSE grade) might be different if they performed differently in different examination series – e.g. a student entering early might have performed poorly initially then improved at a later date, while the converse might be true for a student that entered early and lost the motivation to improve their performance. There were therefore multiple options for modelling the relationship between prior attainment and outcomes to form the basis of the statistical predictions for those students who had entered on multiple occasions in 2012 – 2013. Essentially, the question was, for students that entered on multiple occasions in 2012 – 2013 (of which there were many – see Chapter 6), which of their grades should be used in the basis of the predictions – their best grade, their first grade, or their last grade? The implications of each option are outlined below.

The first option to consider is using a student's best grade. This approach would mean that students in summer 2014 would be rewarded with the same value added as the highest performance of students who had entered on multiple occasions in 2012 – 2013. Since students receive their best grade as their final grade, then this might be considered the fairest option. However, given the findings in Chapter 6 and 7 of this thesis, the use of

a student's best grade in the basis of the predictions is likely to mean that any increase in outcomes due to multiple entry and the unreliability in assessment (rather than improvements in students' performance) would be carried forward via the statistical predictions. This might artificially increase outcomes, not due to the performance of the students, but due to the presence of EMEE in previous academic years (see Chapter 6 for a discussion). This would likely compromise the comparability of standards over time, as the value added relationship would have changed due to EMEE and the error that is inherent in any assessment. Thus, standards would not be comparable in terms of the causes of attainment (see Newton, 2010a).

Given that using a student's best grade therefore seems problematic, an alternative option was to use a student's first grade. This would mean that any increase in grades resulting from EMEE would not be carried forward via the statistical predictions. Using a student's first grade might therefore have been an attractive option politically given the concerns around the possible negative implications of EMEE (see Chapter 2). However, there are considerable issues with this approach.

The main issue is that for students who had previously entered on multiple occasions, their first entry attempt was likely to be an early entry. While some students might have performed to their potential when entering early, this is unlikely to be the case for the majority of students (see DfE, 2011). Students entering early would have had less teaching time relative to those entering at the end of Year 11 and would be less mature (see Chapter 2). Thus, a student's first grade is unlikely to be a true reflection of how student's generally perform at the end of Year 11. As such, basing the summer 2014 predictions on a student's first grade (for multiple entrants) would likely under-predict how students would perform at the end of Year 11 – the very students for whom the summer 2014 predictions would be generated for. This approach would therefore likely disadvantage students certificating in summer 2014 and again would risk the comparability of standards over time given the causal definition outlined previously (see

Newton, 2010a), though the effect would be the opposite of using the best grade: students in summer 2014 would likely be predicted to achieve a lower grade compared to students with the same prior attainment in the years preceding EMEE. Using a student's first grade therefore seems undesirable.

A final option to consider is to use the last grade that a student achieved in the basis of the statistical predictions. In many ways, this might offer the most like with like comparison with the students certificating in summer 2014, since the majority of students were likely to have achieved their last grade in the summer examination series (i.e. the final series that they could enter in Year 11). However, this approach is not entirely unproblematic either. Students in the 2012 – 2013 academic year who entered multiple times might have been less motivated in their last entry attempt than those students entering in summer 2014 as a single entry attempt, particularly if they had achieved or exceeded their target grade prior to this. Furthermore, some students might not have been taught for their final entry attempt. In both of these cases, the value added based on a student's last entry attempt might have been too low. Additionally, for some students their last entry attempt might have been prior to the summer series – e.g. in January or March. Since these students would have had less teaching time and would have been less mature (see Chapter 2), they might have performed differently to those entering in the summer.

To inform the discussion of which approach to use further, consideration was given to the implications that each approach would have for the statistical predictions. Table 27 shows the predicted outcomes for the 2013 – 2014 academic year using either a student's first, best, or last grade (for multiple entrants) in the basis of the predictions (Standards and Technical Issues Group, unpublished). As one might expect from the discussions above, using a student's best grade would yield the highest outcomes, using a student's first grade would yield the lowest outcomes, and using a student's last grade would fall between the two. However, perhaps what is most striking is the size of the differences, particularly at grade C. For example, there is a 9.5% difference in the predicted outcomes

when using a student's first or best grade in the basis of the prediction. While the larger differences around grade C are not surprising given that schools were more likely to enter their borderline grade C students on multiple occasions (see Chapters 5 and 6), it is evident that the decision of which grade to use would have a large impact on the positioning of the grade boundaries and the grades that students achieved. Furthermore, there would be significant implications for the comparability of standards. Thus, the decision on which approach to use was highly important, particularly when it is considered that the largest differences were at grade C, a key threshold grade in GCSE mathematics for accountability measures and students' future prospects.

Table 27. *Predictions for the 2013 – 2014 academic year based on results in the 2012 – 2013 academic year (n = 446,384 in 2013 – 2014)*

	A*	A	B	C	D	E	F	G
Best grade	5.4	16.9	36.7	70.8	82.6	89.6	95.0	98.5
Last grade	5.4	16.5	35.7	68.0	80.9	88.2	93.9	97.9
First grade	4.9	15.3	32.8	61.3	77.2	86.6	93.5	97.8

Given the discussions above and the impact on the predictions, the decision was taken by the awarding bodies to use the last grade that a student achieved in the 2012 – 2013 academic year in the basis of the predictions for summer 2014 (see Ofqual, 2014b, for details of the predictions in summer 2014). This was considered to be the most like-with-like comparison possible, and aimed to ensure that students were appropriately rewarded in summer 2014. Furthermore, this aimed to ensure that standards were comparable, given a student's prior attainment, and therefore satisfy the casual definition of standards discussed previously in this chapter (see Newton, 2010a).

In the event, the outcomes of each awarding bodies awards for GCSE mathematics in summer 2014 met the respective predictions (i.e. those based on the last grade of EMEE students in the previous academic year), suggesting that the approach used had

effectively supported the maintenance of standards.[113] However, it is hoped that the discussions here have demonstrated how shifts in entries can pose challenges for maintaining standards. Thus, while practices such as EMEE can have consequences for schools and students, there are also significant technical considerations. Indeed, had different decisions been taken in summer 2014, then the outcomes would likely have been very different, with significant implications for the comparability of examination standards and the grades that students achieved. The discussions here have focused on EMEE in GCSE mathematics, yet this is not the only case where shifts in entries (resulting from changes to the availability of assessments and policy changes) have posed challenges for the maintenance of standards, GCSE English in 2012 being another notable example (e.g. see Isaacs, 2014, and Ofqual, 2012, for details).

Summary of Chapter 8

To summarise, this chapter has drawn together the previous findings in this thesis by considering the implications of EMEE for examination standards. This has involved considering the theoretical literature relating to how examination standards are defined, how this literature relates to the comparable outcomes approach to maintaining standards that is used in England for GCSEs, and the implications of EMEE for examination standards.

The chapter has focused on the rise in EMEE prior to accountability changes in September 2013, and the subsequent fall in EMEE from November 2013. While these

[113] When awarding grades each year it is not necessarily straightforward to determine whether standards have been maintained. While it is obvious whether predictions have been met or not, it is less obvious when predictions are met but they are unreliable. Issues might therefore only come to light once results are issued to schools, and the extent to which such issues surface might depend on what the results are like. For example, if results are all lower or more variable than schools anticipate then there is likely to be considerable fallout (e.g. as in the case of GCSE English in 2012; see Isaacs, 2014; Ofqual, 2012), but it is debatable the extent to which there would be the same fallout if results were all higher than schools anticipated.

changes in entries have had different implications for examination standards, fundamentally the issue is the same: that when considering examination standards through a causal definition, as defined by Newton (2010a) and implied by comparable outcomes, EMEE poses challenges to how far standards can be considered to be comparable.

The first set of discussions focused on the rise in EMEE. In many respects this poses the greater challenge for ensuring comparability, as it is difficult to conceive how comparability can ever be assured in the presence of EMEE (assuming that the comparable outcomes approach to maintaining standards is adopted). This is because of the error that is inherent in assessment and the possible cumulative effect of students achieving a higher grade than their 'true' score would suggest (see Chapter 6). Thus, while at face value outcomes may seem comparable in a single examination series, this is not the case when students enter multiple times and retain their highest grade as their final grade. Approaches to mitigating this risk have been discussed in this chapter, but none seem satisfactory. Furthermore, while there are other approaches in other jurisdictions that may support comparability in the presence of EMEE (e.g. an equating approach perhaps), this would require fundamental changes to the approach currently used in England (e.g. see Wheadon, 2011), and it is not obvious that this would solve the issue.

The second set of discussions focused on the fall in EMEE following changes to accountability measures announced in September 2013. While there were also technical considerations here to ensuring comparability, these seemed less challenging and in practice the approach adopted seemed to achieve this (although as noted in this chapter, determining the extent to which standards have actually been maintained is not always straightforward). Nonetheless, this example highlights how shifts in entries can pose challenges for maintaining standards, and the need to consider the implications of any changes in entries.

In considering the implications of EMEE for examination standards this chapter has added to the previous literature in a number of ways. Although the discussions have focused on operational activities and considerations at time (e.g. how standards are maintained in practice), the chapter extends beyond this by linking these operational considerations to the theoretical literature relating to how examination standards are defined, an approach that is not frequently adopted. Indeed, much of the debate relating to definitions of examination standards has taken place between different individuals to those involved in the decisions about how standards should be maintained in practice. This is not necessarily problematic, but bringing the issues together, as this chapter has done, is also considered to be important. Being an insider in the assessment system (having been employed by an awarding body and the regulator) means that I am in a fairly unique position to do this.

In addition to providing this link, the chapter has gone further by specifically focusing on the implications of EMEE for examination standards and situating this approach within the theoretical literature, a link that has not previously been made. Indeed, the issue of how EMEE might have implications for examination standards has received very little attention in the literature, even from a more operational perspective, despite the implications that have been outlined in this thesis (see also Chapters 6 and 7).

Chapter 9 – Discussion

Introduction

Within the system of high stakes assessments in England there is an expectation that students will take qualifications at certain points in their education – GCSEs are typically taken aged 16, and A levels aged 18. While this represents the typical and indeed expected route through the assessment system, this thesis explored an alternative approach to entering GCSE mathematics – EMEE. This approach was defined as instances where students entered GCSE mathematics throughout Year 10 or midway through Year 11, and also entered on multiple occasions before the end of Year 11 (i.e. on at least two occasions). EMEE rose in prominence between 2007 and 2013 before changes to accountability measures were announced in September 2013 meaning that, henceforth, only a student's first entry would count in school performance measures. The prevalence of EMEE declined sharply following this announcement and single examination entry at the end of Year 11 once again became the norm. This thesis therefore provides a study of EMEE as the story unfolded.

The thesis initially focused on early entry since when this work was planned much of the concern and public debate focused on this approach to entering GCSE assessments (see Chapter 1). The scope then widened to consider multiple entry as the link between the two approaches became key. The research used a multiphase mixed methods approach comprising three phases and iterating between quantitative and qualitative methods, to consider five over-arching research questions (see Chapter 2; subsidiary research questions are provided in each chapter, as appropriate):

> RQ1) What are the key trends in early entry to GCSE mathematics assessments between 2007 and 2011?
>
> RQ2) What influences schools' entry approaches to GCSE mathematics assessments, in particular, to entering students early and on multiple occasions?

RQ3) What characteristics at the school and student level are associated with multiple entry to GCSE mathematics assessments?

RQ4) What are the implications of EMEE for examination outcomes in GCSE mathematics assessments?

RQ5) What are the implications of EMEE for maintaining examination standards?

Quantitative methods were used to consider the trends in early entry and the possible effects of EMEE for national examination outcomes in GCSE mathematics (see Chapters 4, 6 and 7), while qualitative methods were used to explore the factors that influenced schools' approaches to entering students to GCSE mathematics assessments, focusing on EMEE (see Chapter 5). The findings from all of the phases were then considered in light of the implications for maintaining examination standards (see Chapter 8). This included a consideration of the theoretical literature relating to how examination standards are defined, as well as a consideration of the methods that are used to maintain GCSE standards in practice.

This chapter is structured such that a review of the findings is provided, followed by a discussion of the challenges, limitations and contributions of this research. Finally, future research ideas are discussed and the research is concluded by considering what the research suggests for EMEE.

Review of findings

Increases in early entry to GCSE English and mathematics

The first phase of this research used quantitative methods to consider the key trends in early entry between 2007 and 2011 (see Chapter 4), thus addressing RQ1: what are the key trends in early entry to GCSE mathematics assessments between 2007 and 2011? The analyses focused on specific instances of early entry – students entering early at the end of Year 10 (see Chapter 4) – and considered trends in the number of early entrants

for a number of subjects, before focusing on GCSE English and mathematics where early entry was most prevalent.

The key finding showed an increasing trend for students to enter GCSE English and mathematics early between 2007 and 2011, in particular to the foundation tier of the assessments (see also ACME, 2011; DfE, 2011; Gill, 2010, 2012, 2013; Ofsted, 2013; Ofqual, 2013b). There was also a slight increase in early entry for other large entry subjects that contribute to the EBacc. However, since the proportion of early entries in these other subjects remained relatively small (or in the case of languages was likely to be due to a specific reason, such as the presence of native speakers; see Ofsted, 2013), the remaining analyses focused on English and mathematics.

The increase in early entry for GCSE English and mathematics was coupled with a decline in the achievement of students entering early, as well as a decline in their prior attainment, suggesting that the cohort of students that entered early was changing over time (see also DfE, 2011; Gill 2013). While traditionally, early entry was reserved for the more able students, a wider range of students were increasingly being entered early. This is supported by evidence showing that it was comprehensive schools, secondary modern schools and academies that increasingly entered a greater percentage of their students early, with fewer early entries from independent schools (see also DfE, 2011). Furthermore, there was evidence that early entry was more prevalent in schools with a higher percentage of students eligible for FSM and schools with a lower percentage of students achieving five GCSEs at grades A* – C (including English and mathematics), the key performance measure at the time (see Chapter 1). This resonates with previous findings (see DfE, 2011).

Despite this, there was also evidence that 15-year-old students certificating early performed worse than 16-year-old students certificating at the end of Year 11. This difference tended to be greatest for the lower ability students, which seems at odds with the increase in early entry for weaker students. While this might suggest an effect of early

entry on attainment, it is important to note that these comparisons only focused on the outcomes in one specification at a particular point in time. Thus, if students entered early and re-sat their outcomes might change.

The findings from the first quantitative phase replicate and support the existing literature (see ACME, 2011; DfE, 2011; Gill, 2010, 2012, 2013; Ofsted, 2013; Ofqual, 2013b; Zanini, 2015). They also extend the literature by providing a more focused analysis of trends in early entry over a longer time period than has previously been available (see Chapter 2). For example, the analyses considered changes in the type of students that entered early over a five year period. The analyses also considered additional variables to the existing literature. For example, they considered the trends in early entry by tier of entry to the assessment, suggesting that early entry was more prevalent on the foundation tier than the higher tier. There may have been many reasons for this, some of which were discussed in the qualitative phase of this research (see Chapter 5).

In summary, the first phase of this research suggested that the increase in early entry for GCSE English and mathematics was for a wider range of students than just the most able. This suggested that early entry was not a random occurrence and there were factors influencing schools' entry approaches. In particular, the findings at the school level suggested that one of the drivers of early entry was likely to be the pressure on schools to achieve results. Such pressures might result in decisions being made by individual schools, in terms of their entry policies, to maximise the chance of students achieving a particular grade, most likely a grade C. This was explored in the second (qualitative) phase of this research.

The influence of performance measures on schools' entry approaches

The second phase of this research used qualitative methods to consider the factors that guided schools' approaches to entering students to GCSE mathematics assessments, focusing on EMEE (see Chapter 5). This phase therefore addressed RQ2: what influences schools' entry approaches to GCSE mathematics assessments, in particular, to

entering students early and on multiple occasions? Semi-structured interviews were conducted with 15 individuals based in 14 schools in England, focusing on five key areas:

- entry approaches to GCSE mathematics prior to the accountability changes announced in September 2013, focusing on the use of EMEE;
- entry approaches post-accountability changes;
- the process of deciding entry approaches within the school;
- the influences on the chosen approaches and potential consequences; and
- the effects of entry approaches on teaching and how students are prepared for assessments.

The use of qualitative methods aimed to address some of the limitations of previous research (see Chapters 2 and 3), by allowing a more in-depth analysis of the factors that guide schools' approaches to entering students to GCSE assessments. The interviews were conducted shortly after changes to accountability measures were announced (DfE, 2013a) that effectively ceased widespread EMEE. This allowed respondents to reflect on their approaches to entering students to GCSE mathematics prior to these changes, and consider their approaches going forwards (see Chapter 3 for a discussion of the possible implications of this).

The key finding highlighted how the pressure to achieve results had driven schools' approaches to entering students to assessments, prompting widespread EMEE and associated teaching practices. EMEE was particularly focused on borderline grade C students, potentially to the detriment of other students in the school. The interviews therefore highlighted the direct link between early entry and multiple entry – schools' desire to allow students as many entry attempts as possible before the end of Year 11 meant that they had to enter students early to facilitate this, even if they were not necessarily ready for the assessments.

The entry policies used in schools were highly complex, varying both between and within individual schools for different groups of students, despite the overall desire to achieve results being a key driver. While this desire was frequently associated with the need to improve results for the school, there were also instances of schools wanting to improve results to promote their students' future life chances. As discussed in Chapter 5, it is worth noting that prior to the accountability changes announced in September 2013 the needs of the schools and students were largely aligned, since improving results for the school through EMEE was also beneficial for the students. However, once it was announced that only a student's first entry would count in school performance measures this was not necessarily the case, requiring schools to balance the needs of the school and the needs of the students. For the majority of schools, the needs of the school appeared to take precedence.

The interviews suggested that schools also adapted their entry approaches in response to a number of factors. This included the availability of assessments (which frequently changed for GCSE mathematics at the time the interviews were conducted), the success or otherwise of previous approaches, and external pressures such as performance measures. Thus, when changes to school performance measures were announced in September 2013 the majority of respondents reported that their school had withdrawn their entries from the November 2013 examination series due to the risk that early entry was thought to pose to their final results. This again highlights the clear link between early and multiple entry and the influence of performance measures on schools' approaches.

While early entry was frequently coupled with multiple entry, some schools reported that they used standalone early entry for their higher ability students or what they described as 'at-risk' students – those students who might not remain in education until the end of Year 11 (see also ACME, 2011; Pope and Noyes, 2011). Both of these approaches were likely to continue, as it was perceived to be in the best interests of these students.

In addition to generating evidence about the influences on schools' entry approaches, the findings from the qualitative study also showed that schools carefully considered how they prepared their students for assessments within the context of EMEE. Students were often coached to pass examinations with teaching focused on topics that were (considered) likely to appear on the assessments, or topics where students were (considered) likely to achieve marks. To facilitate this, many schools 'set' students into teaching groups that were targeted at achieving particular grades, frequently moving students between these groups after each examination series. This fluidity allowed schools to focus resources on particular students, typically those on the C/D borderline, and was considered an effective method of increasing the chances that a student would achieve a grade C. While some of these practices are likely to exist in a system of single examination entry, they were facilitated by EMEE and the opportunities that this presented.

The qualitative study also considered some of the possible implications of EMEE. A number of respondents thought that EMEE was likely to have increased the proportion of students achieving the grade C threshold at the national level, not necessarily due to any change in the ability of the students, but because the teaching practices associated with EMEE (e.g. 'teaching to the test' and the cycles of revision and assessments) were thought to artificially inflate outcomes. There were also thought to be implications for students, particularly for their motivation. Some respondents thought that EMEE would reduce motivation since students would become disengaged through repeatedly preparing for and sitting assessments, while other respondents argued that it could motivate students by giving them the confidence to improve their performance. Evidently, EMEE is not a 'one size fits all' approach and the implications are likely to differ for different groups of students and within different schools.

While the qualitative findings suggested that EMEE was often driven by the needs of the schools, it would be remiss not to acknowledge the impact that this was having within schools. Many respondents, of whom the majority were classroom teachers, reported the

moral dilemma that they faced in the context of EMEE: on the one hand they considered it their moral duty to get students a grade C so would use whatever methods necessary to achieve this, yet on the other hand they considered certain entry and teaching practices (enacted for the benefit of the school) as immoral, due to the possible negative implications for students' learning. This tension was evident throughout many of the interviews, leading some respondents to support the changes to accountability measures announced in September 2013 since they thought this would reduce the prevalence of EMEE and afford them with greater teaching time.

In summary, the qualitative study has provided further insight into schools' approaches to EMEE than has previously been available. This helps to explain some of the trends in early entry observed in Chapter 4 (and the existing literature; see Chapter 2), thus supporting the multiphase design of this research. For example, the analyses in Chapter 4 identified that schools with lower performance outcomes were more likely to enter students early. When this trend is considered in light of the qualitative findings it is not particularly surprising, given that the lower performing schools are likely to experience greater pressure to improve results than the higher performing schools. Similarly, the approach of entering the higher ability students early for the foundation tier to 'bank' a grade C, discussed by some schools in the qualitative phase, is likely to be one explanation for the greater increase in early entry to the foundation tier (see Chapter 4).

The findings from the qualitative phase resonate with much previous research describing how the pressure on schools to maximise results influences behaviour, in terms of schools' approaches to both teaching and learning. For example, within the existing literature there is evidence of schools adopting various approaches to teaching their students to maximise results, including the scheduling of intensive revision sessions, setting students by ability, and 'teaching to the test' (Ball, Maguire, Braun, Perryman, and Hoskins, 2012; Burgess, Propper, Slater, and Wilson, 2005; Gillborn and Youdell, 2000; Goldstein, 2004; Hutchings, 2015; Noyes, Drake, Wake, and Murphy, 2010; Perryman,

Ball, Maguire and Braun, 2011; Stobart, 2008; West, 2010; Wilson, Croxson and Atkinson, 2006; Wilson and Piebalga, 2008). The current study adds to this literature by demonstrating how schools have adopted approaches to teaching their students and preparing them for assessments in the context of EMEE.

This research also adds to the literature by suggesting a clear link between the pressure of performance measures and EMEE, evidence that has not previously been available. This adds to the existing literature that shows how schools can adapt their entry approaches to maximise results, particularly at the grade C threshold: there is previous evidence of students being entered for alternative assessments to GCSEs that were rewarded highly in performance tables (Hutchings, 2015; Perryman, Ball, Maguire and Braun, 2011; Stobart, 2008; Wolf, 2011) or perceived as 'easier' (Burgess, Propper, Slater and Wilson, 2005; West, 2010), of students not being entered if they are considered unlikely to achieve the required threshold grade (de Wolf and Janssens, 2007), and of schools strategically entering students to tiered assessments (Gillborn and Youdell, 2000; Strand, 2012). In the case of EMEE, the link with performance measures was particularly clear given the sharp decline in EMEE following changes to accountability measures in September 2013 (see Chapters 5 and 8).

Rising early and multiple entry = rising outcomes?

The third phase of this research used quantitative methods to consider the characteristics of students entering on multiple occasions and the possible implications of EMEE for examination outcomes (see Chapters 6 and 7). This phase therefore addressed RQ3: what characteristics at the school and student level are associated with multiple entry to GCSE mathematics assessments?, and RQ4: what are the implications of EMEE for examination outcomes in GCSE mathematics assessments? This examines one of the issues identified by respondents in the qualitative phase – that is, the possible effect of EMEE on examination outcomes (see Chapter 5). While previous research has focused on the possible implications for individual students (see DfE, 2011; Gill, 2012, 2013, 2014;

Ofsted, 2012, 2013; Vidal Rodeiro, 2014), these analyses focused on the implications for outcomes at the cohort level, though the two are clearly linked.

There were two parts to this phase: the first was an empirical analysis of multiple entry and the implications for examination outcomes using the NPD (see Chapter 6), and the second was a simulation study exploring the possible effects of multiple entry on outcomes (see Chapter 7). The two chapters complement one another by offering a different perspective on the implications of EMEE for examination outcomes. Chapter 6 considers EMEE in practice in the 2012 – 2013 academic year (when it was likely at a peak) by using empirical analyses of national data, while Chapter 7 models the influence of different levels of multiple entry on outcomes.

Using the NPD, the analyses in Chapter 6 showed that for the cohort completing Year 11 in the 2012 – 2013 academic year around half of the students had entered on more than one occasion. While the majority of students entered twice, there was also evidence of students entering on a greater number of occasions (up to twelve times, although this was rare and restricted to a handful of cases). The students that entered on multiple occasions were shown to be weaker in terms of their prior attainment, and they also tended to achieve lower outcomes. This resonates with previous findings in the literature (see DfE, 2013b, c), and provides additional evidence relating to the number of times that multiple entrants certificated.

EMEE was also shown to have possible implications for outcomes. When comparing a student's best grade against the grades from all of their entry attempts the outcomes for the former far exceeded the latter, particularly at grade C. Furthermore, there was evidence that as the number of entry attempts a student made increased, the difference between a student's best grade and their average grade also increased. While this suggests an effect of EMEE on outcomes, it is not possible to disentangle the influence of the various factors that may have contributed to these differences – in particular, to ascertain where outcomes may be legitimately different (due to legitimate changes in

performance), or where they may be increased due to the presence of measurement error in assessment. Thus, the final analysis in Chapter 6 explored the impact of double entry on outcomes when students certificated in the same examination series. This showed that around half of students achieved the same grade on each assessment, replicating previous findings in the literature (see Keeling, 2009; Vidal Rodeiro, 2014). Furthermore, the analyses showed a large effect on outcomes at grade C when implications at the cohort level were considered, an effect that has not previously been considered in the context of EMEE. This effect is arguably less likely to be influenced by levels of test preparedness or motivation since the assessments were sat in the same examination series.

The analysis of multiple entry in Chapter 6 provides evidence relating to the prevalence of multiple entry in the 2012 – 2013 academic year and the characteristics of students that entered on multiple occasions, thus adding to the existing literature (see Chapter 2). Furthermore, the analyses have offered a new perspective on the possible effects of EMEE by considering the implications of EMEE for examination outcomes. While there are a number of caveats to bear in mind (see Chapter 6), the analyses suggest that EMEE is likely to have increased overall outcomes at the cohort level. This could be due to a number of factors including the extent to which students are prepared for the assessments, their motivation and engagement, and the presence of measurement error in assessment. While some of these factors are likely to represent real (and therefore legitimate) changes in performance, others might not, suggesting possible implications for how the grades that students achieve through EMEE should be interpreted – and indeed how national examination outcomes can be interpreted.

To consider this further, simulations were used to model the effect of multiple entry on outcomes when the proportion of students entering on a single or multiple occasion changed (see Chapter 7). This analysis assumed that student factors such as test preparedness and motivation were comparable across sittings, hence focused on the

possible effects of assessment unreliability on outcomes. The results illustrated that, as the proportion of students entering on multiple occasions increased, the difference in outcomes between all students' grades and their best grade also increased (see also Wheadon, 2010). Moreover, as the proportion of multiple entrants increased, examination outcomes for the overall cohort of students also increased, particularly around the grade C threshold. These findings suggest that when the ability of the student cohort and the standards of the qualifications remain constant (as assumed in the simulations), the practice of EMEE itself can increase the outcomes achieved when a student's best grade counts as their final grade. Thus, it is possible that differences in outcomes between two cohorts could reflect the entry approaches of the students, rather than the ability of the cohort. This has implications for how grades and outcomes can be interpreted (depending on how standards are defined), and the process via which standards are maintained.

Implications for maintaining examination standards

The final section of this thesis draws together the previous phases to consider the implications of EMEE for maintaining examination standards (see Chapter 8), thus addressing RQ5) what are the implications of EMEE for maintaining examination standards? The discussions are situated within the theoretical literature relating to how examination standards are defined, and draw on the definitions of standards proposed by Newton (2010a). Based on these definitions, the comparable outcomes approach to maintaining standards used in England for GCSEs since 2011 (see Chapter 1) is considered to imply a causal definition of standards. Thus, standards are considered to be comparable based on the causes of students' attainment – in this case, their prior attainment. Given this definition, there were implications for comparability both when EMEE rose in prominence and when it subsequently declined following changes to accountability measures in September 2013. First, the discussions considered the rise in EMEE.

The comparable outcomes approach to maintaining GCSE standards is operationalised through the use of prior attainment based predictions (see Chapter 1) – thus, if the prior attainment of a cohort remains similar from one year to the next, then outcomes should also remain similar. Given the casual definition of standards, students at key linked boundaries (i.e. grades) would therefore be expected to be comparable in terms of their prior attainment (see Newton, 2010a). However, this definition is challenged in the presence of EMEE. This is because of two (related) points: first, students who enter on multiple occasions always retain their best grade as their final grade, and second, due to EMEE, students are likely to achieve different grades on some entry attempts. The achievement of a different grade might be due to legitimate factors such as further teaching and learning time, but it might also be due to the error that is inherent in assessment. The latter might mean that some students achieve a higher grade on one of their entry attempts that does not reflect the grade that their 'true' score would suggest (see Chapter 6),[114] yet this grade would be their final grade. If this were the case then the relationship between prior attainment and outcomes is likely to be comprised. This challenges our definition of standards, since the causes of attainment are unlikely to be comparable if one were to compare students that entered once with students that entered on multiple occasions. Thus, although comparability might be assured when considering each examination series separately (even in the presence of EMEE), this is unlikely to be the case when students' outcomes across multiple examination series are combined.

The discussions in Chapter 8 therefore focused on possible methods for ensuring comparability in the presence of EMEE, according to our causal definition. These discussions explored the possible policy changes that could facilitate this and logistically whether these are possible, situating the discussions with the context in England at the

[114] Note that other student level factors can also be influential – e.g. motivation and preparedness – but we are concerned here with the factors that do not represent legitimate performance.

time that EMEE was increasing. Possible approaches might include limiting the number of times that students can re-sit a qualification, taking into account the presence of multiple entry when generating evidence to support the setting of grade boundaries, or introducing policies such that only a student's final grade would count. The challenges with each of these approaches are discussed, suggesting that there is no straightforward approach to maintaining standards and therefore ensuring comparability within the presence of EMEE and within the context of comparable outcomes.

This issue became somewhat theoretical since EMEE decreased following changes to accountability measures that were announced in September 2013. However, this change in entry approach also raised challenges for maintaining standards, given our causal definition (Newton, 2010a). This is because the cohort of students changed significantly between summer 2013 and summer 2014 from one that contained a large number of multiple entrants, to one that mainly contained single entrants. This was largely influenced by the policy changes relating to accountability measures, but was also influenced by changes to the structure and availability of the assessments. Thus, Chapter 8 discusses the challenges of this and outlines the approach that was used in practice to mitigate the risk that the decline in EMEE might comprise the comparability of standards given our causal definition (Newton, 2010a).

While both the rise and fall of EMEE had different implications for examination standards, fundamentally the issue is the same: that when considering examination standards through a causal definition, as defined by Newton (2010a) and implied by comparable outcomes, EMEE raises questions about how far standards can be considered to be comparable in terms of students' prior attainment. This suggests that alternative approaches might be necessary to ensure comparability in the presence of EMEE.

Contributions

The work presented in this thesis draws on a number of areas of literature and equally makes a number of contributions. Primarily, it adds to the literature relating to EMEE. At the point that this research was planned such literature was limited and while the literature has subsequently developed (see Chapter 2), there remains a number of gaps that this thesis contributes to. In particular, the thesis provides evidence of the key trends in early entry that extends the previous literature by considering trends over a longer time period and by considering additional variables to the existing literature (for example, tier of entry). Furthermore, it adds to the literature by considering the possible implications of EMEE for examination outcomes at the cohort level, an issue that has not previously been discussed.

This thesis also provides evidence of the influences on schools' approaches to EMEE. The findings of the qualitative phase are considered to be particularly important since at the time this work was conducted much of the debate relating to the possible drivers of EMEE was based on anecdote and not informed by research evidence. Furthermore, there was (and remains) little qualitative evidence exploring the possible drivers of EMEE in England (though more recently evidence has emerged from Wales; see Chapter 2). Thus, the qualitative findings provide greater insight into the influences on schools' approaches to entering students to assessments than has previously been available.

The qualitative phase also adds to the existing literature relating to the way that policy and performance measures in particular can influence and interact with schools' behaviour (see Chapter 2). There is much existing evidence showing how performance and accountability measures can influence schools' teaching approaches, but less relating to schools' approaches to entering students for assessments (though there is some; see Acquah, 2012 for an overview and Chapter 2). Thus, although this thesis does not focus on policy issues specifically, it provides another example of how policy can influence behaviour in schools, by charting the rise and fall of EMEE alongside the relevant policy

decisions relating to performance measures and accountability that influenced this behaviour. It therefore provides additional evidence of the way in which performance measures can influence behaviour within schools that transcends EMEE. This behaviour has been described as 'gaming' (e.g. Goldstein, 2008; Wolf, 2011) and is not unique to the field of education (e.g. see Burgess, Propper, and Wilson, 2005), or the English assessment context (e.g. see Hutchings, 2015, for a summary of findings from the US).

The research has also contributed to the existing literature relating to examination standards. The discussions in Chapter 8 consider the link between the approaches for maintaining standards in practice and the theoretical literature relating to how examination standards are defined, specifically relating to EMEE. These discussions have demonstrated the challenges that EMEE can pose for maintaining standards given our causal definition of standards (Newton, 2010a), an issue that has not previously received attention in the literature. Again, these issues are likely to transcend EMEE and be relevant if and when other shifts in entry approaches occur.

Challenges and limitations

There were a number of challenges and limitations to the overall research programme in this thesis and within each phase of the research (see Chapter 3). While many of the details have been provided in the relevant chapters, the over-arching challenges and limitations are outlined here.

Focus of the research

One of the key aspects of this research is that it focused on EMEE in GCSE mathematics. While there are several reasons for this (see Chapter 1), one obvious limitation is that the findings might not generalise to other subjects. GCSE mathematics is likely to offer the most extreme example of EMEE due to the nature and availability of the assessments (see Chapter 1), suggesting that the situation may differ in other subjects. Thus, while this thesis offers an in-depth account of EMEE in mathematics, it would also have been

interesting to consider EMEE in other subjects (e.g. in both mathematics and English). For some parts of this thesis this would have been possible (indeed Chapter 4 provides some evidence of trends in early entry to English as well as mathematics), but to do so throughout would have been beyond the scope of this thesis, particularly in relation to the qualitative phase. Nonetheless, the drivers and implications of EMEE, and the associated teaching practices, may have been different.

Methodological challenges

Mixed methods approach

Despite the benefits of using a mixed methods approach in this research (see Chapter 3), there are methodological challenges. Many of these were considered at the start of the research so were discussed in Chapter 3, yet some issues can only come to light when the research is nearing completion.

One of the key issues relates to the integration of the data. Mixed methods research can be challenging since the results of each phase can be contradictory (Creswell, 2011; Creswell and Plano Clark, 2011). This was less of an issue in this research since a multiphase design was used that allowed a series of separate studies to consider an overall research topic (Creswell, 2011). Nonetheless, the research must have an overall coherence (Creswell and Plano Clark, 2011). In the current research each phase built on the findings of the previous phase, which helped to ensure that the overall research programme addressed a common research topic. However, this created other challenges: the research was required to adapt as the study progressed and the findings of previous stages were known. This included adapting research questions as the study progressed. This was particularly challenging since the research progressed over a six year period against a backdrop of policy changes (see later discussion in relation to contextual challenges).

Interviews, sampling and recruitment

The challenges of using interviews were discussed in Chapter 3, some of which were experienced in this research, despite attempts to adhere to the principles outlined by Kvale (1996; see Chapter 3). While these issues are not revisited here, there are other limitations of the chosen methodology to consider. First, the qualitative interviews only included the perspectives of some stakeholders in the process. The majority of respondents were classroom teachers and heads of departments, yet it is possible that other themes might have emerged had the views of head teachers, for example, also been sought. Furthermore, the perspective of parents and students might have added alternative perspectives. Some attempt was made to gather information on the possible implications for students – for example, on students' motivation and engagement – but this only related to respondents' perceptions of this, rather than how students actually felt. Discussing the issues with students may have elucidated this further. The perspective of students is often overlooked (see Elwood, 2012), suggesting a possible avenue for further research (see later discussion in this chapter).

Second, there are limitations relating to the sampling and recruitment of schools. Although an 'ideal' approach to sampling schools was devised that aimed to gather the views of a wide range of schools, the approach used necessarily deviated from this due to the challenges of recruiting schools (see Chapter 5). This meant that the final sample might have included a more limited range of views that initially desired – for example, schools that had a certain or strong viewpoint might have been more likely to participate. While the interviews seemed to elicit a range of responses, they are unlikely to be representative of all schools in England and no claims to this effect are made.

Contextual challenges

A further but important issue relates to the contextual challenges in this research. As the research was conducted part-time it necessarily spanned a relatively long time period – around six years. This meant that the research progressed as the story of EMEE

unfolded. While this contributed to the overall approach taken in the research (see Chapter 1), there are contextual challenges. In particular, there have been a number of policy changes that have been announced as the work progressed, some of which have had significant implications for the behaviours being studied (see Chapter 1). This issue is a common reality for those researching in the field of education and educational assessment, due to the pace at which education reform typically occurs (Baird and Lee-Kelley, 2009). However, there are possible implications of this. For example, it might increase the risk that the research becomes irrelevant or obsolete before it is even complete. While this could happen in any research, the risk is likely to be increased when the timescales for the research are longer. To mitigate such risks in the present research the implications of EMEE were considered within a wider context than just the behaviour under study. Thus, the research adds to the literature relating to EMEE, but is also of relevance to other areas such as the influence of performance measures on schools' behaviour, and the theoretical literature relating to how examination standards are defined.

Another challenge due to the timescales of the research is that the literature has to some extent progressed and moved on. Again, this can be particularly challenging when research is conducted over a longer time period. Thus, although some of the analyses presented in the first phase of this research aimed to address gaps in the literature, due to the timescales for completing the research other evidence has since been published that fills some of these gaps. While this research is still considered to make a number of original contributions (see earlier discussion in this chapter), it is acknowledged that some of the analyses now replicate other research that has since been published. This does not mean that the analyses are not useful, since they corroborate previous findings, but it could be argued that they make a more limited contribution to the literature than originally intended.

Further research

It is typical to consider further research in the discussion of any research, but in the present case there are arguably two ways of thinking about this – if EMEE was still rife, what further research would be useful?; and given that EMEE has now declined (in England at least), what further research would be useful? On the surface the former question might seem less relevant since EMEE has declined in England, but as discussed earlier in this thesis, the same issue has been observed in Wales in more recent years (see Chapters 1 and 2). Thus, it is possible that EMEE may be prevalent again in the future, at another time or in another place.

If EMEE were still prevalent further research could usefully explore a number of areas: this includes considering EMEE in different subjects such as GCSE English, where schools' approaches (particularly to teaching) are likely to be different due to the different nature of the subject content and the assessments. Given this, the implications of EMEE for students might also be different – for example, younger students might find engaging with the content of English more challenging than mathematics, due to a lack of maturity (see Chapter 2).

Further research might also consider the other possible implications of EMEE. The focus here has been on the implications for examination outcomes and the maintenance of standards, but there are other issues as well. This includes implications for students, including the impact of EMEE on motivation and engagement, and any subsequent impact on future progression and performance. While there have been a small number of quantitative analyses considering the latter issues, qualitative methods would add further depth and provide additional insight into the implications of EMEE for students. They would also allow the voice of the student to be heard, a perspective that is often lacking (Elwood, 2012). Furthermore, they would allow the implications for different types of students to be considered, since the effects are likely to differ. This might raise issues

around fairness and equity, since there are likely to be positive and negative effects of any approach (e.g. see Gillborn and Youdell, 2000).

Additional research might also consider further the implications of EMEE for examination standards. The discussions in this thesis relating to the implications of EMEE for examination standards were situated within the context of how examination standards were maintained in England at the time that EMEE rose in prominence (i.e. the comparable outcomes approach). Exploring alternative approaches to maintaining standards might therefore be useful. This might also be informative should alternative methods for maintaining standards in GCSE assessments become necessary more generally – for example, if prior attainment measures are no longer available to generate statistical predictions (see Wheadon, 2011).

These suggestions for further research are situated within the context of EMEE still being present, but given that EMEE has now ceased, what other research might be useful? Evidently there is more work to be done in relation to defining examination standards, given some of the issues that have been raised. For example, challenges to the theoretical definitions proposed by Newton (2010a) have been raised (see Baird and Gray, 2016), particularly where the methods for maintaining standards are not considered to align with any of these definitions. In this thesis we have focused on the statistical evidence for maintaining standards since this drives the process through comparable outcomes, but it has also been argued that comparable outcomes implies a phenomenal definition of standards due to the use of examiner judgement (Baird and Gray, 2016). This thinking is currently evolving and will continue to be debated, transcending the issues relating to EMEE discussed here.

The challenges that EMEE pose for maintaining standards through comparable outcomes (and specifically the use of statistical evidence) suggests that further research relating to the use of alternative methods for setting grade boundaries, in the context of GCSEs, might also be beneficial. Researching alternative approaches to harnessing judgemental

evidence of students' performance to support the maintenance of standards is currently on the agenda for Ofqual, the exam board regulator in England, who have committed to this in their corporate plan (Ofqual, 2017e). However, if judgemental evidence came to the fore this might imply an alternative definition of standards based on students' attainment (rather than the causes of their attainment; see Newton, 2010a). Furthermore, it is difficult to conceive how such approaches would be beneficial in the presence of EMEE since many of the issues of using statistical evidence would also apply to using judgemental evidence. Nonetheless, this is a wider issue than just within the context of EMEE (or indeed the English assessment context), and further research considering this would be useful.

Concluding comments

This thesis has focused on the trends, drivers and possible implications of EMEE. So in conclusion, what does this research suggest for EMEE? It is clear from the quantitative analyses that there are trends in EMEE relating to the type of students that entered early and on multiple occasions, with EMEE being more prevalent for lower ability students and students from lower performing schools. This was supported by the findings in the qualitative phase. However, it was also apparent that within the umbrella term of EMEE schools' adopted a number of different approaches to entering students to assessments. Thus, entry approaches were highly complex and individualistic, and this also fed into schools' teaching approaches, which again varied considerably both within and between individuals schools. Furthermore, while the majority of schools reported that they used EMEE to improve results due to the pressures of performance measures, there were instances of schools using EMEE to strategically support their students – for both the more able students and students considered to be at-risk of not completing Key Stage 4.

This suggests that on an individual level EMEE might serve one student well, yet for another it is likely to be detrimental – it is not a 'one size fits all' approach. Thus, drawing

any conclusions about the use of EMEE on an individual level is difficult. However, when we consider the implications of EMEE from a system perspective, the position is perhaps a bit clearer. The evidence from the second quantitative phase, when considered in light of the implications for examination standards, suggests that EMEE might result in some students achieving a grade that does not reflect their 'true' score, thus raising concerns about the comparability of examination standards. From a standards perspective, it is therefore difficult to conceive of EMEE as a good thing: indeed, if standards cannot be maintained in the presence of EMEE, then the currency of the grades that students achieve is undermined. This raises concerns around how those grades can be interpreted, given their importance in allowing students access to further opportunities and in performance measures: if grades are not comparable, then their use is questioned – essentially, the very purpose of awarding grades is undermined.

While the prevalence of EMEE has decreased in England, given the cyclical nature of policy changes and reforms (Baird and Hopfenbeck, 2016), many of the issues identified in this thesis are likely to recur at another time, or indeed in another place, perhaps in relation to alternative approaches to entering students to assessments. It is therefore hoped that the evidence provided here can be informative and contribute to the debate, if or when other changes in entries pose challenges for the maintenance of standards within the context of the comparable outcomes approach in England.

References

Acquah, D. (2013). *School accountability in England: past, present and future*. Manchester, UK: AQA.

Advisory Committee on Mathematics Education (2011). Early and Multiple entry to GCSE Mathematics. Retrieved from http://www.acme-uk.org/news/news-items-repository/2011/5/position-paper-on-early-and-mutiple-entry-to-gcse-mathematics

Al-Bayatti, M.F., and Jones., B. (2003). *Statistical study of the differences in candidates' results between first and second attempts in some GCE AS units*. Paper presented at the British Educational Research Association Annual Conference, 11 – 13 September, Edinburgh, UK.

Alvesson, M., and Skoldberg, K. (2000). *Reflexive Methodology: New Vistas for Qualitative Research*. London: Sage.

Armitage, E. and Lau, C. (2017). *Is the English Baccalaureate a passport to future success?* Manuscript submitted for publication.

Assessment and Qualifications Alliance (2016). Guide to the Uniform mark scale (UMS). Retrieved from http://store.aqa.org.uk/admin/results-days/AQA-UMS-GUIDE.PDF

Association of School and College Leaders (2011). Legitimate reasons for GCSE early entry. Retrieved from http://www.ascl.org.uk/opinion/press_releases/legitimate_reasons_for_gcse_early_entry

Association of School and College Leaders (2013). More piecemeal changes are damaging to students' education. Retrieved from http://www.ascl.org.uk/news-and-views/news_news-detail.html?id=more-piecemeal-changes-are-damaging-students-education

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, and P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards* (pp. 124 – 156). London: Qualifications and Curriculum Authority.

Baird, J., Beguin, A., Black, P., Pollitt, A., and Stanley, G. (2012). The Reliability Programme Final Report of the Technical Advisory Group. In D. Opposs, and Q. He (Eds.), *Ofqual's Reliability Compendium* (pp. 771 – 838). Coventry: Office of Examinations and Qualifications Regulation.

Baird, J., Cresswell, M.J., and Newton, P.E. (2000). Would the real gold standard please step forward? *Research Papers in Education, 15*, 213 – 229.

Baird, J., and Gray, L. (2016). The meaning of curriculum-related examination standards in Scotland and England: a home–international comparison. *Oxford Review of Education*, *42*(3), 266 – 284.

Baird, J., and Hopfenbeck, T.N. (2016). Curriculum in the Twenty-First Century and the Future of Examinations. In D. Wyse, L. Hayward, and J. Pandya (Eds.), *The Sage Handbook of Curriculum, Pedagogy and Assessment Vol. 2* (pp. 821 – 837). London: Sage.

Baird, J., and Lee-Kelley, L. (2009). The dearth of managerialism in implementation of national examinations policy. *Journal of Education Policy, 24*(1), 55 – 81.

Baird, J., and 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.

Baker, M. (2007, January 13). League Tables: Only Half the Story. *BBC News.* Retrieved from http://news.bbc.co.uk/1/hi/education/6257063.stm

Ball, S., Maguire, M., Braun, A., Perryman, J., and Hoskins, K. (2012). Assessment technologies in schools: 'Deliverology' and the 'play of dominations'. *Research Papers in Education, 27*, 513 – 533.

Balls, E. (2008a, June 10). Written Ministerial Statement on the National Challenge for Secondary Schools. Retrieved from http://www.publications.parliament.uk/pa/cm200708/cmhansrd/cm080610/wmstext/80610m0001.htm

Balls, E. (2008b, October 14). Commons speech on National Curriculum tests. Retrieved from http://www.publications.parliament.uk/pa/cm200708/cmhansrd/cm081014/debtext/81014-0004.htm

Bell, J.F. (2001). Patterns of subject uptake and examination entry 1984 – 1997. *Educational Studies, 27*(2), 201 – 219.

Benton, T. (2016). *Comparable Outcomes: Scourge or Scapegoat?* Cambridge: Cambridge Assessment.

Berg, B. L. (2001). *Qualitative Research Methods for the Social Sciences, 4th edn.* London: Allyn and Bacon.

Bond, R., Green, D., and Jaworski, B. (2009). Motivating years 12 and 13 study of mathematics: researching pathways in Year 11. *Proceedings of the British Society for Research into Learning Mathematics*, *29*(3), 19 – 24.

Boyd, W. L. (2000). Editorial: What counts as educational research? *British Journal of Educational Studies, 48*, 347 – 351.

Bradshaw, J., and Wheater, R. (2013). Reporting error and reliability to test-takers: an international review. *Research Papers in Education*, *28*(1), 106 – 117.

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77 – 101.

Braun, A., Maguire, M., and Ball, S.J. (2010). Policy enactments in the UK secondary school: examining policy, practice and school positioning. *Journal of Education Policy*, *25*(4), 547 – 560.

Brennan, R. (2001). An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement, 38*, 295 – 317.

British Educational Research Association (2011). Ethical guidelines for Educational Research. Retrieved from http://content.yudu.com/Library/A2xnp5/Bera/resources/index.htm?referrerUrl=http://free.yudu.com/item/details/2023387/Bera

Bryan, K., and Digby, A. (1983). Pupil performance and school effectiveness. *Westminster Studies in Education*, *6*(1), 33 – 41.

Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative Research, 6*, 97 – 113.

Bryman, A. (2012). *Social research methods*. Oxford: Oxford University Press.

Burgess, S., Propper, C., Slater, H., and Wilson, D. (2005). *Who Wins and Who Loses from School Accountability? The Distribution of Educational Gains in English Secondary Schools*. Bristol: The Centre for Market and Public Organisation.

Burgess, S., Propper, C., & Wilson, D.J. (2005). *Will more choice improve outcomes in education and health care? The evidence from economic research*. Bristol: The Centre for Market and Public Organisation.

Centre for Analysis of Youth Transitions (2011). *The English Baccalaureate and GCSE choices.* London: DfE.

Centre for Education Research and Practice (2013). Research Code of Practice version 2.5. Manchester, UK: AQA.

Chamberlain, S. (2008). Sleeping in or selecting out?: Candidates' absence from GCSE examinations. *Research in Education*, *79*, 53 – 66.

Chamberlain, S. (2013). Qualification users' perceptions and experiences of assessment reliability. *Research Papers in Education, 28*(1), 118 – 133.

Cheek, J. (2005). The practice and politics of funded qualitative research. In N. K. Denzin and Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research 3rd Edition* (pp 387 – 410). Thousand Oaks, CA: Sage.

Chowdry, H., and Sibieta, L. (2011). *Trends in education and schools spending*. London: Institute for Fiscal Studies.

Chyriwsky, M., and Kennard, R. (1997). Attitudes to able children: a survey of mathematics teachers in English secondary schools. *High Ability Studies, 8*(1), 47 – 59.

Cizek, G.J., and Bunch, M.B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests.* Thousand Oaks, CA: Sage.

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

Clauser, A.L., and Wainer, H. (2016). A Tale of Two Tests (and of Two Examinees). *Educational Measurement: Issues and Practice*, *35*(2), 19 – 28.

Coe, R. (2010). Understanding comparability of examination standards. *Research Papers in Education, 25*, 271 – 284.

Colangelo, N., Assouline, S.G., and Gross, M.U.M. (2004). *A nation deceived: How schools hold back America's brightest students (Vols. 1-2).* Iowa City: University of Iowa, The Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Cookson, T. (2012, April 23). The steep climb to University starts with GCSEs. *The Guardian*. Retrieved from http://www.telegraph.co.uk/education/9221229/The-steep-climb-to-university-starts-with-GCSEs.html

Corbin. J., and Morse, J.M. (2003). The unstructured interactive interview: issues of reciprocity and risks when dealing with sensitive topics. *Qualitative inquiry*, *9*(3), 335 – 354.

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

Cresswell, M.J. (2003). *Heaps, prototypes and ethics: The consequences of using judgements of student performance to set examination standards in a time of change*. London: Institute of Education.

Creswell, J.W. (2011). *Educational research: planning, conducting, and evaluating quantitative and qualitative research (Fourth Ed.).* Harlow: Pearson Education.

Creswell, J.W., and Plano Clark, V. (2011). *Designing and Conducting Mixed Methods Research*. Thousand Oaks: Sage.

Cuff, B., Meadows, M., and Black, B. (2017). *An investigation into the Sawtooth Effect in UK secondary school assessments*. Manuscript submitted for publication.

Curtis, P. (2009, January 15). Schools accused of 'hot housing' to get results. *The Guardian.* Retrieved from http://www.guardian.co.uk/education/2009/jan/15/schools-tables-exams

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

de Wolf, I.F., & Janssens, J.G. (2007). Effects and side effects of inspections and accountability in education: An overview of empirical studies. *Oxford Review of Education, 33*, 379 – 396.

Department for Children, Schools and Families (2007). Effective provision for gifted and talented students in secondary education. London: DCSF.

Department for Education (2010). The Importance of Teaching. The Schools White Paper. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/175429/CM-7980.pdf

Department for Education (2011). *Early entry to GCSE examinations*. London: DfE.

Department for Education (2012). Early entry. Letter from Michael Gove to Her Majesty's Chief Inspector. Retrieved from http://www.education.gov.uk/inthenews/inthenews/a00204891/ofsted-asked-to-discourage-gcse-early-entry-a-damaging-trend-harming-interests-of-many-pupils

Department for Education (2013a). Changes to early entry at GCSE. Retrieved from https://www.gov.uk/government/news/changes-to-early-entry-at-gcse

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

Department for Education (2013c). *Statistical first release: GCSE and equivalent results in England 2012/2013 (provisional).* London: DfE.

Department for Education (2015a). Destinations of key stage 4 and key stage 5 students, 2012/13. Retrieved from https://www.gov.uk/government/statistics/destinations-of-key-stage-4-and-key-stage-5-pupils-2012-to-2013

Department for Education (2015b). English Baccalaureate (EBacc) policy paper. Retrieved from https://www.gov.uk/government/news/new-reforms-to-raise-standards-and-improve-behaviour

Department for Education (2017a). Progress 8 and Attainment 8. Guide for maintained secondary schools, academies and free schools. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/583857/Progress_8_school_performance_measure_Jan_17.pdf

Department for Education (2017b). Implementing the English Baccalaureate: Government consultation response. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/630713/Implementing_the_English_Baccalaureate_-_Government_consultation_response.pdf

Department for Education and Employment (1997). Excellence in Schools. London: DfEE.

Department for Education and Skills (2001). Schools Achieving Success. London: DfES.

Department for Education and Skills (2005a). 14-19 Education and Skills. London: DfES.

Department for Education and Skills (2005b). The Schools White Paper: Higher standards, better schools for all. London: DfES.

Department of Education and Science (1988). Education Reform Act 1988. London: DES.

Eason, S. (2010). *Predicting GCSE outcomes based on candidates' prior achieved Key Stage 2 results*. Guildford, UK: AQA.

Eason, S. (2015). *Additional trialling of alternative KS2-based prediction models for GCSE in advance of the summer 2015 exam series*. Guildford, UK: AQA.

Education and Skills Funding Agency (2017). 16-19 funding: maths and English condition of funding. Retrieved from https://www.gov.uk/guidance/16-to-19-funding-maths-and-english-condition-of-funding

Elwood, J. (2012). Qualifications, examinations and assessment: Views and perspectives of students in the 14–19 phase on policy and practice. *Cambridge Journal of Education, 42*, 497 – 512.

Feilzer, M.Y. (2010). Doing Mixed Methods Research Pragmatically: Implications for the Rediscovery of Pragmatism as a Research Paradigm. *Journal of Mixed Methods Research, 4*(1), 6 – 16.

Finlay, L. (2002a). Negotiating the swamp: the opportunity and challenge of reflexivity in research practice. *Qualitative Research*, *2*(2), 209 – 230.

Finlay, L. (2002b). "Outing" the Researcher: The Provenance, Principles and Practice of Reflexivity. *Qualitative Health Research, 12*(4), 531 – 545.

Fontana, A., and Frey, J.H. (2005). The interview: From neutral stance to political involvement. In N. K. Denzin and Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research (3rd Ed.)* (pp. 695 – 727). Thousand Oaks, CA: Sage.

Fox, A., and Allan, J. (2014). Doing reflexivity: moments of unbecoming and becoming. *International Journal of Research and Method in Education*, *37*(1), 101 – 112.

Freeman, J. (2002). *Out-of-school educational provision for the gifted and talented around the World*. London: DfE.

Gardner, J. (2013). The public understanding of error in educational assessment. *Oxford Review of Education, 39*(1), 72 – 92.

Gill, T. (2010). *How old are GCSE candidates? Statistics Report Series No. 20.* Cambridge: Cambridge Assessment.

Gill, T. (2012). *Age distribution of GCSE candidates in England 2011. Statistics Report Series No. 45*. Cambridge: Cambridge Assessment.

Gill, T. (2013). Early entry GCSE candidates: Do they perform to their potential? *Research Matters: A Cambridge Assessment Publication, 16*, 23 – 40.

Gill, T. (2014). *An investigation of the effect of early entry on overall GCSE performance, using a Propensity Score Matching method*. Paper presented at the British Educational Research Association Annual Conference, 23 – 25 September, London: UK.

Gillborn, D., & Youdell, D. (2000). *Rationing education: Policy, practice, reform and equity.* Buckingham: Open University Press.

Goldstein, H. (2004). Education for all: the globalization of learning targets. *Comparative Education, 40*(1), 7 – 14.

Goldstein, H. (2008). Evidence and education policy – some reflections and allegations. *Cambridge Journal of Education, 38*, 393 – 400.

Gorard, S. (2002). The role of secondary data in combining methodological approaches. *Educational Review*, *54*(3), 231 – 237.

Greenbank, P. (2003). The Role of Values in Educational Research: the case for reflexivity. *British Educational Research Journal*, *29*(6), 791 – 801.

Haertel, E.H. (2006). Reliability. In R.L. Brennan (Ed.), *Educational Measurement 4th edition* (pp. 65 – 110). Westport, CT: American Council on Education/Praeger.

Hall, A. (2012, March 8). Statement on early entry GCSE. Retrieved from http://www.aqa.org.uk/supporting-education/news/statement-on-early-entry-gcse

Hanson, B.A. (2008). Wither qualitative/quantitative? Grounds for methodological convergence. *Quality and Quantity, 42,* 97 – 111.

Hanson, B.A., and Brennan, R.L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score models. *Journal of Educational Measurement*, *27*(4), 345 – 359.

Harlen, W. (1994). Developing public understanding of education – a role for researchers. *British Educational Research Journal*, *20*(1), 3 – 16.

He, Q., and Opposs, D. (2012). The reliability of results from national tests, public examinations, and vocational qualifications in England. *Educational Research and Evaluation*, *18*(8), 779 – 799.

He, Q., Opposs, D., and Boyle, A. (2012). A Quantitative Investigation into Public Perceptions of Reliability in Examination Results in England. In D. Opposs and Q. He (Eds.), *Ofqual's Reliability Compendium* (pp. 727–767). Coventry: Office of Qualifications and Examinations Regulation.

Heinrich, M., and Pinot de Moira, A. (2012). *Early examination entry – a first look at 2011 modular GCSE unit outcomes for 15 and 16-year-olds pupils*. Guildford, UK: AQA.

Henderson, S., Holland, J., McGrellis, S., Sharpe, S., and Thomson, R. (2012). Storying qualitative longitudinal research: sequence, voice and motif. *Qualitative Research*, *12*(1), 16 – 34.

Henry, J. (2008, October 18). Taking GCSEs early harms pupils chances of good university places. *Daily Telegraph*. Retrieved from http://www.telegraph.co.uk/education/secondaryeducation/3222199/Taking-GCSEs-early-harms-pupils-chances-of-good-university-place.html

Hodgson, A., and Spours, K. (2011). Rethinking general education in the English upper secondary system. *London Review of Education*, *9*(2), 205 – 216.

Hutchings, M. (2015). *Exam factories? The impact of accountability measures on children and young people*. London: National Union of Teachers.

Independent Academies Association. (2013). Letter to Michael Gove. Retrieved from http://www.iaa.uk.net/wp-content/uploads/2013/10/IAA-Letter-to-Michael-Gove-01-Oct-2013.pdf

Ipsos Mori (2013). Revised: The effects of the English Baccalaureate. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/183528/DFE-RR249R-Report_Revised.pdf

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

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

Jacob, B.A., and Levitt, S.D. (2003). *Rotten apples: An investigation of the prevalence and predictors of teacher cheating*. National Bureau of Economic Research.

Johnson, F., Burrough, N., Ziff, A., and Collins, B. (2012). Public Perceptions of Reliability in Examinations. In D. Opposs and Q. He (Eds.), *Ofqual's Reliability Compendium* (pp. 667 – 689). Coventry: Office of Qualifications and Examinations Regulation.

Johnson, R.B., and Onwuegbuzie, A.J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, *33*(7), 14 – 26.

Johnson, S., and Johnson, R. (2012). Conceptualising and Interpreting Reliability. In D. Opposs and Q. He, *Ofqual's Reliability Compendium* (pp. 459 – 521). Coventry: Office of Qualifications and Examinations Regulation.

Joint Council for Qualifications (2014). GCSE Full Course Results by Age Group 2014. Retrieved from https://www.jcq.org.uk/examination-results/gcses/2014/gcse-full-course-results-by-age-group-2014

Joint Council for Qualifications (2015). GCSE Full Course Results by Age Group 2015. Retrieved from http://www.jcq.org.uk/examination-results/gcses/2015/gcse-full-course-results-by-age-group-2015

Joint Council for Qualifications (2016). GCSE Full Course Results by Age Group 2016. Retrieved from https://www.jcq.org.uk/examination-results/gcses/2016/gcse-full-course-uk-by-age-2016

Joint Council for Qualifications (2017). Instructions for conducting examinations. Retrieved from https://www.jcq.org.uk/exams-office/ice---instructions-for-conducting-examinations/instructions-for-conducting-examinations-2017-2018.

Keeling, H. (2009). *GCSE Maths Double Entries*. Guildford, UK: AQA.

Koshy, V., and Casey, R. (2005). Actualizing mathematical promise: possible contributing factors. *Gifted Education International, 20*, 293 – 305.

Kounali, D., Robinson, T., Goldstein, H. and Lauder, H. (2008). *The probity of free school meals as a proxy measure for disadvantage.* Centre for Multilevel Modelling Research Article. University of Bristol: Bristol.

Kvale, S. (1983). The qualitative research interview – a phenomenological and a hermeneutical mode of understanding. *Journal of Phenomenological Psychology*, *14*(2), 171 – 196.

Kvale, S. (1994). Ten standard objections to qualitative research. *Journal of Phenomenological Psychology*, *25*(2), 147 – 173.

Kvale, S. (1996). *Interviews. An Introduction to Qualitative Research Interviewing*. London: Sage.

Kvale, S. (2006). Dominance Through Interviews and Dialogues. *Qualitative Inquiry, 12,* 480 – 500.

Lee, W., Hanson, B.A., and Brennan, R.L. (2002). Estimating Consistency and Accuracy Indices for Multiple Classifications. *Applied Psychological Measurement*, *26*(4), 412 – 432.

Leech, N.L., and Onwuegbuzie, A.J. (2007). An Array of Qualitative Data Analysis Tools: A Call for Data Analysis Triangulation. *School Psychology Quarterly*, *22*(4), 557 – 584.

Leech, N.L., and Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality and Quantity: International Journal of Methodology, 43*, 265 – 275.

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

Lipsett, A. (2008, October 14). Unions welcome end of secondary Sats. *The Guardian.* Retrieved from http://www.guardian.co.uk/education/2008/oct/14/sats-scrapped-reaction

Livingston, S.A., and Lewis, C. (1995). Estimating the Consistency and Accuracy of Classifications Based on Test Scores. *Journal of Educational Measurement*, *32*(2), 179 – 197.

Lord, F.M., and Novick, M.R. (1968). *Statistical theories of mental test scores*. Oxford: Addison-Wesley.

Lord, F.M., and Wingersky, M. (1984). Comparison of IRT true-score and equipercentile observed-score equatings. *Applied Psychological Measurement*, *8*, 452 – 461.

Mansell, W. (2007, August 24). Schools risk becoming 'exam factories'. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/news/uknews/1561167/Schools-risk-becoming-exam-factories.html

Mansell, W. (2011, January 11). The English BAC causes fury in schools. *The Guardian*. Retrieved from http://www.guardian.co.uk/education/2011/jan/11/league-table-secondary-english-baccalaureate.

Mathematics in Education and Industry (2009). *The effects of 2-tier GCSE Mathematics on transition to AS and A level*. Trowbridge, UK: MEI.

Mauthner, N.S., and Doucet, A. (2003). Reflexive accounts and accounts of reflexivity in qualitative data analysis. *Sociology*, *37*(3), 413 – 431.

Mauthner, N.S., Parry, O., and Backett-Milburn, K. (1998). "The data are out there, or are they?" Implications for archiving and revisiting qualitative data. *Sociology*, *32*(4), 733 – 745.

Maxwell, J.A. (2005). *Qualitative research design: An interactive approach*. London: Sage.

Miles, M.B., and Huberman, A.M. (1994). *Qualitative data analysis: An expanded sourcebook*. London: Sage.

Millman, J. (1989). If at first you don't succeed: setting passing scores when more than one attempt is permitted. *Educational Researcher*, *18*(6), 5 – 9.

Morgan, D.L. (1996). Focus Groups. *Annual Review of Sociology*, *22*, 129 – 152.

Morgan, D.L. (2001). Focus Group Interviewing. In J.F.Gubrium and J.A. Holstein (Eds.), *Handbook of Interview Research – Context and Method* (pp. 141 – 160). Thousand Oaks, CA: Sage.

Murphy, R. (2004). Grades of uncertainty: reviewing the use and misuses of examination results. Retrieved from https://www.atl.org.uk/Images/Grades%20of%20uncertainy.pdf

Murphy. R. (2013). Media roles in influencing the public understanding of educational assessment issues. *Oxford Review of Education*, *39*(1), 139 – 150.

Newton, P.E. (2005a). The public understanding of measurement inaccuracy. *British Educational Research Journal*, *31*(4), 419 – 442.

Newton, P.E. (2005b). Examination standards and the limits of linking. *Assessment in Education: Principles, Policy and Practice, 12*, 105 – 123.

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

Newton, P.E. (2010a). Contrasting conceptions of comparability. *Research Papers in Education*, *25*(3), 285 – 292.

Newton, P.E. (2010b). Thinking About Linking. *Measurement: Interdisciplinary Research and Perspectives*, *8*(1), 38 – 56.

Newton, P.E. (2011). A level pass rates and the enduring myth of norm referencing. *Research Matters: A Cambridge Assessment Publication*, *Special Issue 2*, 20 – 26.

Newton, P.E. (2013). Ofqual's Reliability Programme: a case study exploring the potential to improve public understanding and confidence. *Oxford Review of Education*, *39*(1), 93 – 113.

Newton, P.E., Baird, J., Goldstein, H., Patrick, H., and Tymms, P. (2007). *Techniques for monitoring the comparability of examination.* London: Qualifications and Curriculum Authority.

Node, P., Rutt, S., Schagen, S., and West, A. (2007). *Evaluation of the two year Key Stage 3 project*. London: DfE.

Noyes, A. (2009). Exploring social patterns of participation in university-entrance level mathematics in England. *Research in Mathematics Education*, *11*(2), 167 – 183.

Noyes, A., Drake, P., Wake, G., and Murphy, R. (2010). *Evaluating mathematics pathways: Final report*. London: DfE.

Noyes, A., and Sealey, P. (2011). Managing learning trajectories: the case of 14-19 mathematics. *Educational Review, 63*(2), 179 – 193.

Noyes, A., Wake, G., and Drake, P. (2011). Widening and increasing post-16 mathematics participation: pathways, pedagogies and politics. *International Journal of Science and Mathematics Education*, *9*, 483 – 501.

Ofsted (2012). *Mathematics: made to measure. Messages from inspection evidence*. Manchester: Ofsted.

Ofsted (2013). *Schools' use of early entry to GCSE examinations: its usage and impact.* Manchester: Ofsted.

Ofsted (2014). School inspection handbook: Handbook for inspecting schools in England under section 5 of the Education for Act 2005. Manchester: Ofsted.

Ofsted (2017). HMCI's commentary: October 2017. Retrieved from https://www.gov.uk/government/speeches/hmcis-commentary-october-2017

Ofqual (2009). GCSEs the official student guide to the system. Retrieved from http://webarchive.nationalarchives.gov.uk/20141031163546/http://www.ofqual.gov.uk/files/GCSE_Guide.pdf

Ofqual (2011). *Annual Qualifications Market Report*. Coventry, UK: Ofqual

Ofqual (2012). *GCSE English 2012*. Coventry, UK: Ofqual

Ofqual (2013a). GCSE reform consultation. Retrieved from http://webarchive.nationalarchives.gov.uk/20141031163546/http:/www.ofqual.gov.uk/files/2013-06-11-gcse-reform-consultation-june-2013.pdf

Ofqual (2013b). *Annex C. Early entry in summer 2012 and summer 2013 GCSEs and IGCSEs.* Coventry, UK: Ofqual

Ofqual (2013c). *Annex B. Multiple entry in GCSE/IGCSE English/English language and mathematics in summer 2012*. Coventry, UK: Ofqual

Ofqual (2014a). Su*mmer GCSE and IGCSE entries for England: Provisional Figures April 2014*. Coventry, UK: Ofqual

Ofqual (2014b). Summer 2014 Data Exchange Procedures – Version 2. GCSE, GCSE and Level 1/2 Certificates. Retrieved from http://webarchive.nationalarchives.gov.uk/20141031180321/http://ofqual.gov.uk/standards/summer-2014-exams/

Ofqual (2015a). Corporate Plan 2015-2018. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/419223/2015-03-30-ofqual-corporate-plan-2015-18.pdf

Ofqual (2015b). Setting GCSE, AS and A level standards in summer 2014 and 2015. Retrieved from 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 (2015c). Comparability of Different GCSE and A Level Subjects in England: An Introduction: ISC Working Paper 1. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/606041/1-comparability-of-different-gcse-and-a-level-subjects-in-england-an-introduction.pdf

Ofqual (2017a). Statement from the qualification regulators on changes to GCSEs, AS and A levels. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/635579/Statement_from_the_qualification_regulators_on_changes_to_GCSEs_AS_and_A_levels_2017_update.pdf

Ofqual (2017b). Reviewing GCE and GCSE outcome data received from awarding organisations as part of the data exchange procedure summer 2017. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/620338/Reviewing_outcome_data.pdf

Ofqual (2017c). Summer 2017 GCSE, AS and A level exams A summary of our monitoring. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/639880/Summer_2017_monitoring_summary.pdf

Ofqual (2017d). Provisional summer 2017 exam entries: GCSEs, AS and A levels. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/619387/Summer_2017_exam_entries_GCSEs_AS_and_A_levels_in_England.pdf

Ofqual (2017e). Corporate Plan 2017 – 2020. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/633421/Ofqual-17-6174-Corporate_Plan_2017-20-27.07.17.pdf

Ofqual (unpublished). Provisional and final entries to the November 2013 examination series.

Patai, D. (1994). (Response). When method becomes power. In A. Gitlen, (Ed.), *Power and Method* (pp. 61-73), New York: Routledge.

Paton, G. (2010, August 23). Universities use GCSEs to 'filter out' weak pupils. *The Telegraph.* Retrieved from http://www.telegraph.co.uk/education/educationnews/7960643/Universities-use-GCSEs-to-filter-out-weak-pupils.html

Paton, G. (2012, May 9). Schools forced to spend £328m on 'overinflated' exams. *The Telegraph.* Retrieved from http://www.telegraph.co.uk/education/educationnews/9254585/Schools-forced-to-spend-328m-on-over-inflated-exams.html

Paton, G. (2013, Aug 18). School standards threatened by early and often GCSEs. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/education/educationnews/10250548/School-standards-threatened-by-early-and-often-GCSEs.html

Paton, G., and Moore, M. (2008, October 15). Sats for 14-year-olds abolished: Teachers and parents praise decision. *The Telegraph.* Retrieved from http://www.telegraph.co.uk/education/primaryeducation/3199156/Sats-for-14-year-olds-abolished-Teachers-and-parents-praise-decision.html

Pell, G., Boursicot, K., and Roberts, T. (2009). The trouble with resits… *Assessment and Evaluation in Higher Education*, *34*(2), 243 – 251.

Perryman, J., Ball, S., Maguire, M., and Braun, A. (2011). Life in the pressure cooker – school league tables and English and mathematics teachers' responses to accountability in a results driven era. *British Journal of Educational Studies*, *59*(2), 179 – 195.

Perryman, J., Maguire, M., Braun, A., and Ball, S. (2017). Surveillance, Governmentality and moving the goalposts: The influence of Ofsted on the work of schools in a post-panoptic era. *British Journal of Educational Studies*. Advance online publication https://doi.org/10.1080/00071005.2017.1372560

Pillow, W. (2003). Confession, catharsis, or cure? Rethinking the uses of reflexivity as methodological power in qualitative research. *Qualitative Studies in Education*, *16*(2), 175 – 196.

Pinot de Moira, A. (2008). *The Effects of Maturity: evidence from Linear GCSE Specifications.* Presentation given at the Ofqual Technical Seminar.

Poland, B.D. (2001). Transcription Quality. In J.F. Gubrium and J.A. Holstein (Eds.), *Handbook of Interview Research – Context and Method* (pp. 629 – 649).Thousand Oaks, CA: Sage.

Poon Scott, E. (2012). Short-term gain at long-term cost? How resit policy can affect student learning. *Assessment in Education: Principles, Policy and Practice, 19*(4), 431 – 449.

Pope, S., and Noyes, A. (2011). Early entry in GCSE Mathematics*. Proceedings of the British Society for Research into Learning Mathematics*, *31*(1), 127 – 131.

Putwain, D.W. (2008). Supporting assessment stress in key stage 4 students. *Educational Studies*, *34*(2), 83 – 95.

Putwain, D.W., and Remedios, R. (2014). The Scare Tactic: Do Fear Appeals Predict Motivation and Exam Scores? *School Psychology Quarterly*, *29*(4), 503 – 516.

QCA (2008). GCSE qualification criteria. For first teaching from September 2009. London: QCA.

QCA (2009). Changes to GCSE Mathematics. London: QCA.

QSR International Pty Ltd. (2012). NVivo qualitative data analysis software, Version 10.

Qualifications Wales (2017). *Approaches to early and multiple entry in Wales*. Newport, Wales: Qualifications Wales.

R Development Core Team. (2014). R: A Language and Environment for Statistical Computing, http://www.R-project.org.

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests.* Copenhagen: Danish Institute for Educational Research.

Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics, 92*, 1394 – 1415.

Reis, S.M., and Renzulli, J.S. (2010). Is there still a need for gifted education? An examination of current research. *Learning and Individual Differences*, *20*(4), 308 – 317.

Richardson, H. (2011, May 10). Half of schools see budget cuts, survey suggests. *BBC News*. Retrieved from http://www.bbc.co.uk/news/education-13346238

Ricketts, C. (2010). A new look at resits: are they simply a second chance? *Assessment and Evaluation in Higher Education*, *35(*4), 351 – 356.

Roberts, L., and Bingham, J. (2010, August 25). GCSE results: Girl, five, claims record while seven-year-old boy receives A*. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/education/educationnews/7962546/GCSE-results-Girl-five-claims-record-while-seven-year-old-boy-receives-A.html

Roulston, K., deMarrais, K., and Lewis, J.B. (2003). Learning to Interview in the Social Sciences. *Qualitative Inquiry, 9*, 643 – 668.

Sandelowski, M. (2003). Tables or Tableaux? The challenges of writing and reading mixed methods studies. In A. Tashakkori and C. Teddlie (Eds.), *Handbook of Mixed methods in Social and Behavioural Research* (pp. 321 – 350). Thousand Oaks, CA: Sage.

Scharaschkin, A., and Baird, J. (2000). The effects of consistency of performance on A level examiners' judgements of standards. *British Educational Research Journa*l, *26*, 343 – 357.

Sellgren, K. (2012, August 23). Anger over 'harsh' English GCSE grades. *BBC News*. retrieved from http://www.bbc.co.uk/news/education-19355956

Sellgren, K. (2013, August 13). Warning to schools over multiple exam entries. *BBC News*. Retrieved from http://www.bbc.co.uk/news/education-23687883

Silverman, D. (2011). *Interpreting qualitative data - 4th edition*. London: Sage.

Smith, E. (2008). Pitfalls and promises: the use of secondary data analysis in educational research. *British Journal of Educational Studies*, *56*(3), 323 – 339.

Spalding, V. (2009). *GCSE Science A: The size and effect of 'If at first you don't succeed, try, try again'*. Guildford, UK: AQA.

Spearman, C.E. (1904a). Proof and measurement of association between two things. *American Journal of Psychology, 15,* 72 – 101.

Spearman, C.E. (1904b). 'General intelligence' objectively determined and measured. *American Journal of Psychology*, *15,* 201 – 293.

Standards and Technical Issues Group (unpublished). Awarding GCSE English/English language, GCSE mathematics and GCSE science in summer 2014. Appendix B: GCSE mathematics (excluding the linked pair).

Stobart, G. (2008). *Testing Times. The uses and abuses of assessment*. Abingdon: Routledge.

Stobart, G., Bibby, T. and Goldstein, H. (2005). *Moving to two-tier GCSE Mathematics examinations. An independent evaluation of the 2005 GCSE Pilot and Trial.* London: University of London Institute of Education.

Stobart, G., and Eggen, T. (2012). High-stakes testing – value, fairness and consequences. *Assessment in Education: Principles, Policy and Practice*, *19*(1), 1 – 6.

Stockford, I. (2012a). *AQA award of GCSE Mathematics B (4365) (Linear specification) - summer 2012.* Manchester, UK: AQA.

Stockford, I. (2012b). *AQA award of GCSE Mathematics (4360) – summer 2012.* Manchester, UK: AQA.

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.

Strand, S. (2012). The White-Black Caribbean achievement gap: Tests, tiers and teacher expectations. *British Educational Research Journal, 38*, 75 – 101.

Stringer, N. (2012). Setting and maintaining GCSE and GCE grading standards: the case for contextualised cohort-referencing. *Research Papers in Education*, *27*(5), 535 – 554.

Sykes, E.D.A., Bell, J.F., and Vidal Rodeiro, C.L. (2009). *Birthdate effects: A review of the literature from 1990-onwards.* Cambridge: Cambridge Assessment.

Tashakkori, A., and Teddlie, C. (2003). *Handbook of Mixed Methods in Social & Behavioural Research.* Thousand Oaks, CA: Sage.

Taverner, S., and Wright, M. (1997). Why go modular? A review of modular A-level Mathematics. *Educational Research, 39*(1), 104 – 112.

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.

Teddlie, C., and Sammons, P. (2010). Applications of mixed methods to the field of educational effectiveness research. In B. P. M. Creemers, L. Kyriakides, and P. Sammons (Eds.), *Methodological advances in educational effectiveness research* (pp. 115 – 152). London: Routledge.

Teddlie, C., and Tashakkori, A. (2003). Major issues and controversies in the use mixed methods in the social and behavioural sciences. In A. Tashakkori, and C. Teddlie, C. (Eds.), *Handbook of Mixed Methods in Social and Behavioural Research* (pp. 3 – 50). Thousand Oaks, CA: Sage.

The Institute of Mathematics and its Applications and the London Mathematical Society (2011). Early entry to GCSE mathematics – not always the right path. Retrieved from http://www.lms.ac.uk/content/press-releases#2011

The National Centre for Excellence in the Teaching of Mathematics (2011). Early entry GCSE - concern grows about the impact on mathematics education. Retrieved from www.ncetm.org.uk/news/31970

The Mathematical Association (2010). On the subject of Early and Repeated Presentation for GCSE Mathematics… Retrieved from http://www.mei.org.uk/index.php?page=newsandnewsid=236

Thomson, P. and Gunter, H. (2011). Inside, outside, upside down: the fluidity of academic researcher 'identity' in working with/in school. *International Journal of Research and Method in Education*, *34*(1), 17 – 30.

Tymms, P., and Fitz-Gibbon, C. (2001). Standards, achievement and educational performance: A cause for celebration? In J. Furlong and R. Phillips (Eds.), *Education, reform and the state: Twenty-five years of politics, policy and practice* (pp. 156 – 173). London: Routledge Falmer.

Vahasantanen, K., and Saarinen, J. (2012). The power dance in the research interview: manifesting power and powerlessness. *Qualitative Research*, *13*(5), 493 – 510.

Vallerand, R.J., Pelletier, L.G., Blais, M.R., Brière, N.M., Senécal, C., and Vallières, E.F. (1992). The academic motivation scale: A measure of intrinsic, extrinsic and amotivation in education. *Educational and Psychological Measuremen*t, *52*, 1003 – 1017.

Vasagar, J. (2012, August 22). English GCSEs marked down to curb grade inflation, say teachers. *The Guardian*. Retrieved from http://www.theguardian.com/education/2012/aug/22/english-gcses-marked-down-teachers

Vidal Rodeiro, C.L. (2009). *Some issues on the uptake of Modern Foreign Languages at GCSE. Statistics Report Series No. 10*. Cambridge: Cambridge Assessment.

Vidal Rodeiro, C.L. (2014). *Multiple entries in GCSE/IGCSE qualifications*. Cambridge: Cambridge Assessment.

Vidal Rodeiro, C.L., and Nádas, R. (2012). Effects of modularity, certification session and re-sits on examination performance. *Assessment in Education: Principles, Policy and Practice, 19*(4), 411 – 430.

Walford, G. (2001). *Doing Qualitative Educational Research – A personal guide to the research process*. London: Continuum.

Warmington, P., and Murphy, R. (2004). Could do better? Media depictions of UK educational assessment results. *Journal of Education Policy*, *19*(3), 285 – 299.

Warmington, P., and Murphy, R. (2007). 'Read All About It!' UK news media coverage of A level results. *Policy Futures in Education*, *5*(1), 70 – 83.

Welsh Government (2017). Changes to early entry exams announced. Retrieved from http://gov.wales/newsroom/educationandskills/2017/changes-to-early-entry-exams-announced/?lang=en

West, A. (2010). High stakes testing, accountability, incentives and consequences in English schools. *Policy and Politics, 38*(1), 23 – 39.

West, A., Mattei, P., and Roberts, J. (2011). Accountability and Sanctions in English Schools. *British Journal of Educational Studies*, *59*(1), 41 – 62.

West, A., and Pennell, H. (2000). Publishing School Examination Results in England: Incentives and consequences. *Educational Studies*, *26*(4), 423 – 436.

Westminster Institute of Education (2006). Briefing document on acceleration. Oxford: Oxford Brookes University.

Wheadon, C. (2010). *The impact of re-sitting on the fairness of grading*. Guildford, UK: AQA.

Wheadon, C. (2011). *An Item Response Theory Approach to the Maintenance of Standards in Public Examinations in England*. Durham theses, Durham University. Available at Durham E-Theses Online http://etheses.dur.ac.uk/615/

Wheadon, C., and Stockford, I. (2010). *Classification accuracy and consistency in GCSE and A level examinations offered by the Assessment and Qualifications Alliance (AQA) November 2008 to June 2009*. Coventry: Office of Qualifications and Examinations Regulation.

Wightwick, A. (2017, 16 October). Early entry GCSEs to change as teachers and education chiefs split over children's best interests. *Wales online*. Retrieved from http://www.walesonline.co.uk/news/education/early-entry-gcses-change-teachers-13764196

Wiliam, D. (2001). Reliability, validity, and all that jazz. *Education 3 – 13*, *29*(3), 17 – 21.

Wilson, D., Croxson, B., and Atkinson, A. (2006). What Gets Measured Gets Done: Headteachers' Responses to the English Secondary School Performance Management System. *Policy Studies, 27*(2)*,* 153 – 171.

Wilson, D., and Piebalga, A. (2008). Performance measures, ranking and parental choice: an analysis of the English school league tables. *International Public Management Journal, 11*(3), 344 – 366.

Winter, J. (2001). (Ed.) (Proceedings of BCME5) *Proceedings of the British Society for Research into Learning Mathematics, 21*(2), 20 – 25.

Winkley, J., and Cresswell, M. (2012). Introduction to the concept of reliability. In D. Opposs and Q. He (Eds.), *Ofqual's Reliability Compendium* (pp. 727 – 767). Coventry: Office of Qualifications and Examinations Regulation.

Wolf, A. (2011). *Review of Vocational Education – The Wolf Report*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/180504/DFE-00031-2011.pdf

Zanini, N. (2015). *Age distribution of GCSE candidates in England in 2013. Statistics Report Series No. 88.* Cambridge: Cambridge Assessment.

Appendix A – overview of key policies

The following timeline provides a summary of the key policies relating to GCSE assessments that are relevant to this thesis. These policies primarily relate to the assessment system in England.

Date	Description
1988	GCSEs first introduced to replace O levels and CSEs, intended to be sat by the majority of the 16-year-old cohort
1992	School performance measures introduced by the Conservative Government of the time, focusing on five GCSE A* – C measure
2007	School performance measures revised so that English and mathematics become a requirement in the five GCSE A* – C measure
2007	GCSE qualification and subject criteria revised – new GCSEs developed, mainly modular qualifications
2008	First award of two tier mathematics GCSE
2009	First award of two tier mathematics GCSE without coursework
2009	Terminal rule introduced for GCSEs so that 40% of the assessment has to be sat in the series that students certificate in
2009	First teaching of revised modular GCSEs (except English, mathematics, and ICT)
2010	First teaching of revised GCSEs in English, mathematics and ICT – including a modular and linear GCSE in mathematics
2010	Publication of 'The Importance of Teaching – The Schools White Paper', signalling a return to linear assessment at GCSE

2011 Comparable outcomes approach introduced by Ofqual for maintaining standards at GCSE

2011 First award of revised GCSEs in all subjects except English, mathematics and ICT

2011 Publication of 'Review of Vocational Education: the Wolf report'

2012 First award of revised GCSEs in English, mathematics and ICT – including modular and linear GCSEs in mathematics

2012 'High value' vocational qualifications announced, following Wolf Review – qualifications that are included in performance measures are limited

2012 Government announced plans to develop English Baccalaureate Certificates (EBCs) to replace GCSEs, with each subject being offered by one exam board

2013 Government announced that existing GCSEs would be reformed – plans for EBCs and single exam board for each subject scrapped

2013 Final January and March examination series for GCSE mathematics

2013 Requirement that all students aged 17 – 18 must be in some form of education, training or employment comes into effect

2013 Government announced that only a student's first GCSE entry will count toward school performance measures (September) – students still receive their highest grade as their final grade

2013 Government announced Progress and Attainment 8 as the key performance measure for all schools from 2016

2014 November examination series for GCSE English and mathematics becomes a re-sit only opportunity

2015 Introduction of Progress and Attainment 8 accountability measures for 'opt-in' schools in England

2015 First teaching of reformed GCSEs in English language, English literature and mathematics, all linear assessment

2016 First award of revised GCSE mathematics in Wales (November), with a significant number of early entries

2016 Introduction of Progress and Attainment 8 accountability measure for all schools in England

2017 First award of revised GCSEs (graded 9-1) in England in English language, English literature and mathematics

Appendix B – development of GCSEs

Since their inception, GCSEs have been through various reforms, with each round of development involving various bodies. The Department for Education (DfE) is responsible for the national curriculum which sets out the programme of study that all students in local-authority maintained schools in England must follow. Ofqual (the exam board regulator) is then responsible for over-seeing the development of the qualifications by awarding bodies (also known as exam boards). In the most recent round of GCSE reform (for qualifications first taught in 2015 for some subjects), this was achieved through an accreditation process. Ofqual reviewed the subject content developed by the DfE and published qualification and subject level conditions that awarding bodies used to develop qualifications. Exam boards then submitted the qualifications they developed to Ofqual, who decided whether the qualifications met the appropriate regulatory requirements. This was achieved through review by a panel that comprised subject experts and Ofqual staff (who reviewed the submission from a regulatory framework perspective), and supported by technical experts from Ofqual. The exam boards, regulated by Ofqual, are then responsible for delivering the qualifications.

Appendix C – marking and moderation of GCSE assessments

Externally assessed examinations

The mode of assessment has implications for how the assessments are marked. Examination papers are usually externally marked by examiners that are employed by the exam boards. Examiners are recruited based on a number of criteria, including their subject knowledge and teaching experience, and are allocated to mark a particular question paper. Examiners are trained to mark each paper (via a process known as standardisation) and are monitored throughout the marking period by more senior examiners to ensure quality of marking.

Non-examination assessment

Non-examination assessment, previously known as coursework or controlled assessment, is generally internally assessed within schools. Teachers mark each students' work and a sample is then moderated by the exam boards, to ensure that each school is applying the correct standard. If a moderator finds that the marks are out of line with the required standard, an adjustment is usually applied to all of the students' work to bring the marks into line. If it is not possible to make an adjustment, the moderators' marks are typically used instead.

Appendix D – generating statistical predictions

Prior attainment based predictions model the relationship between prior attainment and outcomes in a reference year, then use this relationship to predict the expected outcomes of students in the current year. At GCSE, prior attainment is measured as students' mean KS2 score. There are two main steps to generating predictions for each subject, as follows.

First, an 'outcome matrix' is generated for the reference year (see figure D1 for a fictitious outcome matrix). Students in the reference year that are matched to their prior attainment are divided into octiles based on their prior attainment at KS2. A matrix is then created that shows how each octile went on to perform in each GCSE. The top octile (numbered 1 in Figure D1) would include the most able students (based on mean KS2 score), and the bottom octile would include the least able students. Once constructed, this matrix shows the probability of achieving a given grade for students in each octile.

		GCSE grade								
		A*	A	B	C	D	E	F	G	U
KS2 octile	1	70	75	80	85	90	95	100	100	100
	2	65	70	75	80	85	90	95	100	100
	3	60	65	70	75	80	85	90	95	100
	4	55	60	65	70	75	80	85	90	100
	5	50	55	60	65	70	75	80	85	100
	6	45	50	55	60	65	70	75	80	100
	7	40	45	50	55	60	65	70	75	100
	8	35	40	45	50	55	60	65	70	100

Figure D1. *Example outcome matrix*

Using this outcome matrix, it is then possible to predict how students in the current year are expected to perform, given their own prior attainment. For example, using the outcome matrix above, 70% of students in octile 1 might be expected to get a grade A*, 75% might be expected to get an A* or A; 80% might be expected to get an A*, A or B;

and so on. This is repeated for each octile and then aggregated together to form a prediction for the probability of achieving each grade.

The outcome matrix is based on all exam board data from a previous examination series (the reference series). Each exam board then applies the same relationship between prior attainment and outcomes to their current cohort of students. Given that the predictions reflect the prior attainment profile of the students entering for each board, one board might have a higher prediction than another if the prior attainment profile of the students is higher.

Appendix E – early entry trends for EBacc subjects

Table E1: *GCSE Geography – percentage of entries by age group (n = total entries)*

Age	2007	2008	2009	2010	2011
<=14	0.08	0.10	0.13	0.28	0.62
15	0.90	1.13	1.45	1.96	2.54
16	98.03	97.73	97.37	96.77	95.85
17	0.94	1.00	1.00	0.96	0.95
18	0.04	0.03	0.04	0.03	0.03
>=19	0.01	0.01	0.01	0.02	0.01
n	181,162	174,505	168,779	168,938	161,874

Table E2: *GCSE History – percentage of entries by age group (n = total entries)*

Age	2007	2008	2009	2010	2011
<=14	0.02	0.02	0.14	0.34	0.73
15	0.96	1.12	1.36	1.73	2.60
16	98.23	98.08	97.68	97.18	95.96
17	0.76	0.74	0.76	0.71	0.68
18	0.02	0.02	0.02	0.02	0.02
>=19	0.01	0.02	0.03	0.03	0.02
n	200,092	201,034	195,789	198,052	197,027

Table E3: *GCSE French – percentage of entries by age group (n = total entries)*

Age	2007	2008	2009	2010	2011
<=14	3.29	4.41	4.60	4.66	4.10
15	4.89	5.48	6.00	5.57	5.30
16	91.04	89.31	88.55	89.04	89.89
17	0.69	0.71	0.73	0.64	0.64
18	0.03	0.03	0.03	0.02	0.02
>=19	0.06	0.06	0.09	0.07	0.04
n	191,560	179,826	169,216	159,762	139,082

Table E4: *GCSE Spanish – percentage of entries by age group (n = total entries)*

Age	2007	2008	2009	2010	2011
<=14	2.94	2.85	3.97	3.73	3.94
15	3.49	3.78	4.28	5.00	4.37
16	91.41	91.48	89.81	89.59	90.40
17	1.12	1.12	1.14	0.97	0.88
18	0.32	0.26	0.24	0.17	0.11
>=19	0.72	0.52	0.57	0.55	0.30
n	55,058	58,470	59,201	60,339	59,184

Table E5: *GCSE Chemistry – percentage of entries by age group (n = total entries)*

Age	2007	2008	2009	2010	2011
<=14	0.03	0.02	0.04	0.03	0.03
15	1.59	0.99	0.94	0.76	0.86
16	96.60	97.75	97.83	98.24	98.14
17	1.60	1.13	1.07	0.88	0.89
18	0.13	0.07	0.07	0.05	0.04
>=19	0.06	0.04	0.05	0.04	0.04
n	50,879	67,860	82,653	112,704	132,053

Table E6: *GCSE Biology – percentage of entries by age group (n = total entries)*

Age	2007	2008	2009	2010	2011
<=14	0.14	0.09	0.08	0.08	0.18
15	1.82	1.12	1.01	0.83	1.14
16	95.37	97.18	97.40	97.84	97.50
17	2.03	1.26	1.14	0.98	0.95
18	0.21	0.11	0.11	0.06	0.07
>=19	0.41	0.23	0.27	0.21	0.16
n	53,909	73,381	87,704	116,237	134,784

Table E7: *GCSE Physics – percentage of entries by age group (n = total entries)*

Age	2007	2008	2009	2010	2011
<=14	0.03	0.08	0.03	0.03	0.02
15	1.41	1.06	0.97	0.74	0.81
16	96.84	97.61	97.80	98.26	98.26
17	1.54	1.12	1.07	0.88	0.84
18	0.11	0.08	0.08	0.05	0.04
>=19	0.07	0.04	0.05	0.04	0.04
n	50,332	66,913	81,904	111,668	130,835

Table E8: *GCSE Science – percentage of entries by age group (n = total entries)*

Age	2007	2008	2009	2010	2011
<=14	0.22	0.10	0.13	0.38	0.76
15	9.46	8.93	8.00	7.67	8.70
16	89.08	89.92	90.55	90.64	89.06
17	1.08	0.88	1.09	1.03	1.10
18	0.10	0.07	0.11	0.10	0.14
>=19	0.07	0.09	0.11	0.18	0.25
n	547,858	527,736	483,813	434,866	384,339

Table E9: *GCSE Geography – cumulative percentage outcomes for 15-year-olds*

Year	*	A	B	C	D	E	F	G	n
2007	11.49	38.26	60.51	78.97	89.79	94.80	96.58	97.98	1,636
2008	15.05	36.64	55.40	73.69	85.71	90.47	93.11	95.24	1,973
2009	8.42	27.58	48.22	70.78	84.76	91.42	95.22	97.26	2,447
2010	5.90	25.11	43.75	65.26	83.21	90.95	95.22	97.52	3,305
2011	0.95	23.43	43.07	65.34	81.82	91.30	96.18	98.64	4,114

Table E10: *GCSE History – cumulative percentage outcomes for 15-year-olds*

Year	*	A	B	C	D	E	F	G	n
2007	6.37	35.59	57.10	73.43	86.59	93.68	96.71	98.75	1,916
2008	14.04	36.38	57.69	74.94	87.20	93.89	96.75	98.44	2,243
2009	5.56	30.37	48.65	67.04	81.83	90.58	95.12	98.05	2,664
2010	10.07	29.31	48.70	66.13	80.23	88.06	93.72	97.11	3,425
2011	0.76	23.91	43.16	62.68	78.74	88.00	93.91	97.25	5,123

Table E11: *Mean scores by age and KS2 score – AQA English A foundation tier*

KS2 category	15-year-old mean	16-year-old mean	Difference
<=2.67	208.91	227.90	18.99
3.00	217.47	239.90	22.43
3.33	233.76	251.99	18.23
3.67	253.89	265.04	11.15
4.00	269.38	276.83	7.45

Table E12: *Mean scores by age and KS2 score – AQA English A higher tier*

KS2 category	15-year-old mean	16-year-old mean	Difference
3.67	324.68	344.16	19.48
4.00	338.62	354.95	16.33
4.33	355.79	368.92	13.13
4.67	372.20	385.23	13.03
5.00	400.15	414.14	13.99

Table E13: *Mean scores by age and KS2 score – AQA Mathematics A foundation tier*

KS2 category	15-year-old mean	16-year-old mean	Difference
<=2.67	52.77	69.87	17.10
3.00	66.31	81.23	14.92
3.33	75.14	90.65	15.51
3.67	91.22	105.28	14.06
4.00	112.18	120.33	8.15

Table E14: *Mean scores by age and KS2 score – AQA Mathematics A higher tier*

KS2 category	15-year-old mean	16-year-old mean	Difference
4.00	63.03	72.67	9.64
4.33	68.80	81.52	12.72
4.67	81.03	100.72	19.69
5.00	110.99	126.29	15.30

Table E15: *Mean scores by age and KS2 score – AQA Mathematics B foundation tier*

KS2 category	15-year-old mean	16-year-old mean	Difference
<=2.67	170.12	198.07	27.95
3.00	190.26	227.24	36.98
3.33	219.82	254.52	34.70
3.67	259.41	288.69	29.28
4.00	309.03	331.07	22.04

Table E16: *Mean scores by age and KS2 score – AQA Mathematics B higher tier*

KS2 category	15-year-old mean	16-year-old mean	Difference
3.67	364.55	398.66	34.11
4.00	381.59	403.66	22.07
4.33	401.30	419.04	17.74
4.67	428.22	444.40	16.18
5.00	476.98	488.23	11.25

Table E17: *Regression output predicting total scores*

Subject	Variable	*B*	*SE B*	*β*
English A – foundation	Age	63.53	5.14	0.42*
	KS2 score	261.55	22.20	2.05*
	Age * KS2 score	-14.04	1.40	-1.77*
R^2 = 0.10, * p < 0.01				
English A – higher	Age	29.38	4.78	0.15*
	KS2 score	108.83	16.70	0.86*
	Age * KS2 score	-3.31	1.05	-0.42*
R^2 = 0.20, * p < 0.01				
Mathematics A – foundation	Age	45.14	6.95	0.60*
	KS2 score	185.81	29.51	2.08*
	Age * KS2 score	-9.10	1.87	-1.63*
R^2 = 0.24, * p < 0.01				
Mathematics A – higher	Age	10.17	15.15	0.08
	KS2 score	38.01	51.04	0.36
	Age * KS2 score	1.04	3.21	0.16
R^2 = 0.28, * p < 0.01				
Mathematics B – foundation	Age	85.74	11.85	0.40*
	KS2 score	360.23	50.28	1.66*
	Age * KS2 score	-15.87	3.17	-1.18*
R^2 = 0.26 , * p < 0.01				
Mathematics B – higher	Age	98.53	12.13	0.47*
	KS2 score	371.32	40.70	1.98*
	Age * KS2 score	-17.88	2.56	-1.53*
R^2 = 0.22, * p < 0.01				

Appendix F – qualitative paper published in Oxford Review of Education

This is an Accepted Manuscript of an article published by Taylor & Francis in Oxford Review of Education on 19 July 2016, available online:

http://www.tandfonline.com/doi/full/10.1080/03054985.2016.1197829



Oxford Review of Education

ISSN: 0305-4985 (Print) 1465-3915 (Online) Journal homepage: http://www.tandfonline.com/loi/core20

The effects of accountability measures in English secondary schools: early and multiple entry to GCSE Mathematics assessments

Rachel Charlotte Taylor

To cite this article: Rachel Charlotte Taylor (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

To link to this article: https://doi.org/10.1080/03054985.2016.1197829

Published online: 19 Jul 2016.

Submit your article to this journal

Article views: 299

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=core20

Download by: [Ofqual] **Date:** 04 January 2018, At: 03:22

OXFORD REVIEW OF EDUCATION, 2016
VOL. 42, NO. 6, 629–645
http://dx.doi.org/10.1080/03054985.2016.1197829

Routledge
Taylor & Francis Group

The effects of accountability measures in English secondary schools: early and multiple entry to GCSE Mathematics assessments

Rachel Charlotte Taylor

Centre for Education Research and Practice, Assessment and Qualifications Alliance (AQA), Manchester, UK

ABSTRACT

This study used semi-structured interviews with staff in 14 secondary schools in England to explore strategies for entering students to GCSE Mathematics, focusing on the use of early and multiple examination entry. The key findings suggest that teachers were keenly aware of performativity pressures and that this prompted the use of a number of strategic entry practices, including the frequent use of early entry to allow students multiple examination attempts. This approach aimed to maximise the number of students achieving a grade C, the threshold for school accountability measures and for accessing further education and employment opportunities. Widespread use of early and multiple entry is likely to cease in England following amendments to accountability measures announced in 2013, yet this study adds a further perspective to the well-established literature detailing the impact of accountability measures in schools, as well as depicting teachers' views on the effects of early and multiple entry upon pupils' education.

KEYWORDS

Early entry; multiple entry; accountability; performance tables; GCSE; Mathematics

Downloaded by [Ofqual] at 03:22 04 January 2018

Introduction

This paper explores the effects of accountability measures on behaviour in secondary schools in England, focusing on the strategic entry of students to General Certificate of Secondary Education (GCSE) assessments—high stakes qualifications taken at age 16. Specifically, the paper focuses on the use of early and multiple examination entry, behaviours that deviate from the typical and expected route through GCSE assessments. The paper sets out the policy context for the research by considering the introduction of performance 'league' tables and the pressures that this has brought within schools, before defining and discussing the behaviours under study.

Policy context

The policy backdrop to this research centres upon the accountability regime in English secondary schools, a system that stems from the 1988 Education Reform Act (Department of Education and Science [DES], 1988). Of primary relevance to this paper is the introduction

CONTACT Rachel Charlotte Taylor rachel.taylor@ofqual.gov.uk

© 2016 Informa UK Limited, trading as Taylor & Francis Group

of school performance 'league' tables for accountability purposes. First introduced in 1992 by the then Conservative Government, performance tables aimed to hold schools accountable to key stakeholders and provide parents with information about each school's performance (West, 2010; West, Mattei, & Roberts, 2011; West & Pennell, 2000), the key measure being the percentage of students achieving five or more 'good' GCSE passes (grades A*–C).

The five A*–C measure continued to be prioritised in later White Papers (Department for Education and Skills [DfES], 2001) and was strengthened in 2007 with the forced inclusion of GCSE English and Mathematics (DfES, 2005).[1] This sought to ensure that schools were not rewarded for entering students to alternative qualifications that contributed a high tariff towards the five A*–C measure (e.g. see Perryman, Ball, Maguire, & Braun, 2011; Wilson, Croxson, & Atkinson, 2006), yet also elevated the importance of attainment in GCSE English and Mathematics (Perryman et al., 2011)—subjects where a 'good' GCSE pass (Department for Education [DfE], 2010) has traditionally been required to allow students access to further education and employment opportunities.

Performance tables have increasingly become high stakes (Perryman et al., 2011; West, 2010), a position reinforced in 2004 by the introduction of 'floor standards'—targets that specify the minimum percentage of students within each school that must achieve the five A*–C threshold (now including English and Mathematics). Failure to achieve this measure, which has become ever-more stringent, results in schools falling 'below the floor' (DfE, 2013a)—an unenviable position that can result in increased monitoring, inspections, and, ultimately, closure.

The nature of performance tables has resulted in grade C becoming an important threshold at GCSE, increasing the pressure on schools to ensure that their students achieve this, particularly in key subjects such as English and Mathematics (see Gillborn & Youdell, 2000; Noyes & Sealey, 2011; Perryman et al., 2011; Pope & Noyes, 2011). This is considered to have prompted the use of a number of strategies within schools to improve results, including pedagogical approaches such as focusing resources upon borderline students, the scheduling of intensive revision sessions, setting students by ability, and 'teaching to the test' (Ball, Maguire, Braun, Perryman, & Hoskins, 2012; Burgess, Propper, Slater, & Wilson, 2005; Gillborn & Youdell, 2000; Goldstein, 2004; Hutchings, 2015; Noyes, Drake, Wake, & Murphy, 2010; Perryman et al., 2011; Stobart, 2008; West, 2010; Wilson et al., 2006; Wilson & Piebalga, 2008). Such behaviours are not restricted to the English assessment system (see Jacob & Levitt, 2003, and Reback, 2008, for evidence from the US), and are unlikely to be without consequence: there is evidence that focusing on borderline students can be detrimental to the attainment of other students (Burgess, Propper, Slater, et al., 2005; Gillborn & Youdell, 2000), that 'teaching to the test' can distort the curriculum (Goldstein, 2004), and that focusing on particular subjects can narrow the curriculum that students experience (Hutchings, 2015).

Alongside teaching behaviours, the focus on the grade C threshold is thought to have influenced schools' entry strategies to assessments. There is evidence of students being strategically entered for alternative assessments that are rewarded highly in performance tables (Hutchings, 2015; Perryman et al., 2011; Stobart, 2008; Wolf, 2011) or perceived as 'easier' (Burgess, Propper, Slater, et al., 2005; West, 2010), of students not being entered if they are considered unlikely to achieve the required threshold (de Wolf & Janssens, 2007), and of schools strategically entering students to tiered assessments[2] (Gillborn & Youdell, 2000; Strand, 2012). More recently, a further strategy—early and multiple entry—has come to the fore at GCSE, and it is to this strategy that this paper now turns.

Downloaded by [Ofqual] at 03:22 04 January 2018

Early and multiple GCSE examination entry

Students are expected to certificate (or complete) their GCSE qualifications once at the end of Year 11 (age 16). As such, early entry can be defined as any situation where students certificate before the end of Year 11, and multiple entry as any situation where students certificate on more than one occasion. Although these practices are defined separately, they are intrinsically linked: early entry allows students multiple opportunities to certificate before the end of Year 11, when their results will count towards performance league tables.[3]

Traditionally, early entry has been used to accelerate or enrich the progress of the most able students (Bryan & Digby, 1983; see also Bell, 2001; Pope & Noyes, 2011), yet more recently there is evidence of early entry being used for a wider range of students (Advisory Committee on Mathematics Education [ACME], 2011; DfE, 2011). This increase has been coupled with a decline in the grade outcomes achieved by early entrants (DfE, 2011), suggesting that early entry is no longer reserved for the most able. This has led to concerns that early entry could have negative implications for students' achievement, future progression, and learning when it is used for whole school cohorts, rather than selectively as a means of acceleration or enrichment (ACME, 2011; DfE, 2011; Qualifications and Curriculum Authority [QCA], 2009; The National Strategies, 2009).

The increase in early entry has been most notable in GCSE English and Mathematics (DfE, 2011; Gill, 2010; Pope & Noyes, 2011). These subjects are key for performance measures, prompting concerns that schools are entering students early to allow multiple entry attempts (or as described in the grey literature, multiple 'bites of the cherry'), to maximise results (e.g. ACME, 2011). This premise is supported by several sources in the grey literature. With reference to GCSE Mathematics, ACME (2011) believe that, 'an all-consuming focus on league tables is skewing behaviour in schools'—a view shared by The Mathematical Association (2010), who cited government policy and the resultant fears of head teachers, teachers, and parents as a key driver for early (and multiple) entry. Furthermore, there is evidence that the lower-attaining schools, for whom accountability measures are likely to be of more concern (Ball et al., 2012; Hutchings, 2015), have a greater tendency to enter students early (ACME, 2011; DfE, 2011; Noyes et al., 2010).

These concerns led to the reform of accountability measures in September 2013 such that only a student's first entry attempt would count in performance tables (DfE, 2013b). A subsequent drop in early and multiple entry ensued (Ofqual, 2014), offering further support for the proposed link between early and multiple entry and performance measures. Despite this, there are few published papers evidencing accountability measures as the key driver in increasing early and multiple entry (e.g. see Ball et al., 2012; Perryman et al., 2011). Indeed, there is little published research focusing on early and multiple entry more generally. Furthermore, in a study conducted by Ofsted[4] (2013) exploring the use of early entry, accountability measures were not identified as influential in guiding schools' entry strategies. Instead, the desires to raise standards and improve student motivation were reported as key factors. This report canvassed the views of a relatively large sample (113 schools), yet there were limitations. The questionnaire methodology yielded a high proportion of 'other' responses for questions concerning which students were entered early and the factors that influenced the decision to enter early, suggesting that the questions, or the available responses, were inadequately constructed. Furthermore, key terms in the questionnaire were ambiguous. For example, the desire to 'raise standards', cited as one of the drivers of

Downloaded by [Ofqual] at 03:22 04 January 2018

early entry, could refer to improving teaching and learning practices, or, as suggested by Ball et al. (2012), to raising examination outcomes. In addition, it is debatable how candid schools might have been in their responses, given that Ofsted administered the survey.

Considering this, the present study aims to extend our understanding of the way that early and multiple entry have been used by schools. In doing so, it adds to the literature on early and multiple entry, and the literature exploring the impact of accountability measures on behaviours in schools. The study focuses on GCSE Mathematics, one of the key subjects for performance measures where early and multiple entry have been particularly prominent (DfE, 2011). Whilst early and multiple entry have declined since policy changes in September 2013 (DfE, 2013b), they are likely to represent one way in which schools respond when faced with extreme accountability pressures. Stobart (2008) describes the 'continuous game of regulatory cat-and-mouse' that ensues when schools attempt to meet targets while loopholes in the system are closed around them. Thus, the decline in early and multiple entry will likely be replaced by alternative strategies with similar drivers and consequences in the future. Additionally, the findings from this study are likely to be relevant in other contexts or times where these strategic behaviours are possible and have associated rewards.

Downloaded by [Ofqual] at 03:22 04 January 2018

Aims and rationale

This study aimed to explore school-level entry strategies to GCSE Mathematics, focusing specifically upon the use of early and multiple entry and the drivers behind these strategies. Given the paucity of previous research, this study was exploratory and qualitative methods were preferred (Miles & Huberman, 1994). Semi-structured interviews were used since they allow in-depth discussions to be conducted with individual participants (Cohen, Manion, & Morrison, 2000). This was considered important since the way in which schools navigated policy and responded to accountability measures was likely to differ. Further, since the topic was likely to be highly sensitive, potentially requiring participants to reveal behaviours that could be considered unethical, focus groups were not considered appropriate. The use of interviews necessarily limits the sample size, however, meaning that it cannot be assumed that the findings are generalisable.

The study focused on entry strategies for GCSE Mathematics since this subject has experienced considerable flux in entries over recent years (DfE, 2011), is a key subject in performance tables, and arguably lends itself more readily to early and multiple entry than other subjects. GCSE Mathematics is entirely externally assessed, multiple examination series have been available each year, the content between specifications is similar, and multiple specifications with each awarding body[5] have been available for a number of years. Given this, Mathematics could be considered a worst case scenario, suggesting that generalising the findings to other subjects might not be appropriate.

Method

Participants

Data were collected via 14 interviews with individuals based in schools in England; 13 one-to-one interviews and one interview with two participants. This resulted in a total of 15 participants, including four heads of department for GCSE Mathematics, nine teachers of

Table 1. Final sample of schools.

School type	% early entry June 2012	% five GCSEs A*–C	% free school meals
Comprehensive A	98.67	50	30%≤ and <40%
Academy A	23.40	70	10%≤ and <20%
Academy B	*	48	10%≤ and <20%
Academy C	100.00	47	20%≤ and <30%
Academy D	*	56	10%≤ and <20%
Academy E	11.69	56	10%≤ and <20%
Academy F	*	62	10%≤ and <20%
Comprehensive B	5.11	43	20%≤ and <30%
Academy G	30.54	57	10%≤ and <20%
Academy H	0.74	58	10%≤ and <20%
Selective A	0.00	96	<10%
Comprehensive C	26.67	68	10%≤ and <20%
Academy J	0.47	62	<10%
Academy K	2.82	73	<10%

*Entry figures were not available for these schools.

Downloaded by [Ofqual] at 03:22 04 January 2018

GCSE Mathematics, and two examinations officers. In addition to their roles within the school, three participants were employed by one of the UK awarding organisations as either an assistant or senior examiner. Whilst this might have resulted in some participants having more extensive knowledge of the assessment system than others, this was not considered problematic; indeed, it may have facilitated more in-depth discussions or provided an additional perspective. The final sample consisted of 10 male and five female participants.

To ensure that schools using a range of entry strategies were included in the research, participants were sampled based on the percentage of students entered early for GCSE Mathematics in June 2012, the performance outcomes of the school, and the percentage of students eligible for free schools meals. The latter two characteristics were considered important, since early entry has been shown to vary on these dimensions (DfE, 2011). Using these criteria, two lists of schools were generated—those that entered a low (below 25%) or high (above 90%) percentage of their students early for GCSE Mathematics. Initial recruitment used these lists to contact schools by telephone in the areas local to the two researchers that conducted the interviews (the North West and Surrey), ensuring that schools with a range of performance outcomes and percentage of students eligible for free school meals were included. However, since recruitment via this method proved challenging, additional approaches were necessary. First, opportunistic sampling was used through an existing contact in the Mathematics community, who identified participants through a network of Mathematics teachers based in the Midlands. Second, a snowball sampling method was employed, where existing participants recommended colleagues at other schools to approach. The final sample comprised 14 schools (Table 1) including three comprehensive schools, one selective school, and ten academies.[6]

Procedure

Data collection took place between December 2013 and February 2014. Interviews were conducted face-to-face by the author (12 interviews) or an assistant researcher (two interviews). Two pilot interviews were conducted, following which minor amendments were made to the interview schedule. Since this mainly comprised re-ordering the sections in the schedule, rather than making any significant changes to the questions, these data were included in the analyses.

Interviews were conducted in the participants' schools and lasted between 25 and 45 minutes. Prior to each interview, the meaning of key terms used in the interviews was briefly discussed. Early entry was used to refer to instances where a student certificated prior to the end of Year 11 (regardless of when or whether they re-sat at a later date), whilst multiple entry was used to refer to situations where students entered more than once—either by using repeated entry (i.e. early entry followed by a re-sit in a separate examination series), or by entering multiple awarding bodies or specifications in the same examination series. Following this, each interview adhered to a semi-structured schedule. Of primary interest were the entry strategies used for GCSE Mathematics prior to, and after, the accountability changes announced in September 2013, and the drivers behind these strategies. Open-ended questions were used at the start of each section to initiate discussions, followed by prompts and follow-up questions when necessary. Digressions from the interview schedule allowed exploration of interesting points raised by the interviewee. All participants consented to the interviews being recorded and were given a £20 voucher in return for participation.

Downloaded by [Ofqual] at 03:22 04 January 2018

Analysis

Interviews were transcribed verbatim by a transcription company. Primarily, the analysis relied on the transcripts, with the audio recordings used where necessary to clarify meaning or to ascertain the tone of a particular comment. Transcriptions were available shortly after each interview, allowing the main researcher to interweave the data collection and analysis stages of the research and thus use knowledge of the existing data to enhance later data collection, through exploring topics in more detail (Miles & Huberman, 1994).

Data were analysed using dominant theme analysis (Maxwell, 2005), a technique that offers a flexible approach to the analysis of qualitative data (Braun & Clarke, 2006). An inductive approach was used, where the coding of the data proceeded 'without trying to fit it into a pre-existing coding frame or the researcher's analytical preconceptions' (Braun & Clarke, 2006). Nevertheless, as stakeholders in the assessment system, the researchers had prior knowledge or preconceptions which may have influenced the interpretation and reporting of the data. To minimise any potential bias during the analysis the data were analysed by two researchers: the author who had conducted the majority of the interviews (12 interviews) and an assistant researcher who joined the research at the analysis stage.

The analysis comprised three main stages, allowing the researchers to reflect on the data throughout. During the initial stage, both researchers independently read the transcripts in-depth (line by line), whilst making notes of possible codes in the data and emerging ideas for organising the data. Following this, the researchers met to discuss the codes and agree on themes to describe the data. The main researcher then read the transcripts a second time, where instances of support and challenge to each theme were identified. A final discussion of the coding was undertaken between the two researchers. Any disagreements were discussed in-depth and agreement sought.

Findings

The interviews generated a large corpus of data yielding multiple themes. These related to the drivers behind schools' behaviours, the strategies that schools used, and the potential

implications of these. Due to the large volume of data, this paper focuses on the themes relating to schools' entry strategies. This includes the impact of accountability measures on schools' behaviour, the entry strategies that schools used, the flexible and experimental entry approach adopted by schools, the effect on teachers, and the issue of responsibility. The remaining themes not discussed in this paper relate to the teaching strategies adopted within schools and the consequences of these behaviours for students. This includes the way in which students are trained for assessments and set into teaching groups, and the effect on a student's learning, motivation, and engagement.

The results are presented as the key themes that emerged from the data, with supporting quotes from respondents. Although the themes are presented separately, there were several links; indeed the first and dominant theme, relating to the way in which accountability and the pressure to achieve results influences behaviour in schools, underlies many of the other themes.

The grade C culture

The key theme emerging from the data related to how the pressures of results and accountability measures influenced behaviour in schools. These pressures led to schools becoming preoccupied with attempts to maximise their examination results and the adoption of a wide range of strategies to achieve this, potentially at the expense of supporting effective teaching and learning. In particular, the pressures around grade C were evident, leading many schools to strategically focus on this threshold and those students around it. This resonates with much previous research (e.g. Ball et al., 2012; Gillborn & Youdell, 2000; Perryman et al., 2011) and is not surprising given the importance of accountability measures and the potential consequences should schools, and students, not achieve the required results:

> … you have to think of the context of what you're living in and that was if we didn't hit that floor target, we opened a whole load of hell for ourselves. We opened doors for more Ofsted, more inspections, more LEA inspections and your day to day life was going to be hard …. (Head of Department, Academy J)

> … it was like get your C grades or risk shutting. (Teacher, Comprehensive A)

Although results were a key concern for all respondents, the degree to which schools felt under pressure differed, with the lower-performing schools tending to report greater pressure to achieve grades than their higher-achieving counterparts (see also Ball et al., 2012; Hutchings, 2015). Thus, the interviewee from a high-achieving selective school reported few concerns about achieving targets, likely due to the high calibre of their student intake. This should not imply that they did not strive to achieve the best results possible; rather, that the pressures from accountability measures were diminished:

> For me, I believe that what's driven the entry strategy in this school is the pressure from the Government and the LEA on the results of our school. If we didn't step the game up, we would have been put into special measures …. (Teacher, Comprehensive A [low-achieving school])

> … I mean, to be honest, we could put all the kids in at this school early probably at the end of Year 10 and they'd all get reasonable results. (Teacher, Selective A [high-achieving school])

For the majority of schools, the focus on results manifested itself in a focus on the grade C threshold, one of the key measures in school performance tables (DfE, 2010). This often led to an 'exam factory' approach (Mansell, 2007; see also Hutchings, 2015), with schools

Downloaded by [Ofqual] at 03:22 04 January 2018

aiming to get as many students over the grade C threshold as possible, regardless, to a certain extent, of how this was achieved and the implications for teaching and learning (see also Ball et al., 2012; Gillborn & Youdell, 2000; Perryman et al., 2011). Students were often discussed as numbers or percentages, an approach facilitated by there being multiple examination series available throughout the academic year,[7] allowing schools to monitor how many students had crossed the grade C threshold as the year progressed. This resonates with the 'A–C economy' described in earlier work by Gillborn and Youdell (2000):

> If they get a C, excellent, that's one more off the list. (Teacher, Comprehensive A)

> … that helped this school as a whole to see exactly what their statistics were in terms of 'ok, now we've got these exam results we know we have got 63% in the bank, our target is 70%, therefore we need another 7%, this equates to 20 more pupils or something so who are these 20 going to be?'. (Teacher, Academy G)

This pressure, particularly at grade C, was highlighted by the language that respondents used throughout the interviews, with many interviewees reporting the 'need' for students to achieve a particular grade to reach school targets:

> … we're measured on how many Cs we got, it didn't matter if they weren't Bs or As and A*s, so you needed to get Cs because if you didn't get enough Cs, you were classed as being a rubbish school …. (Head of Department, Academy J)

Whilst for many schools the drive to achieve as many grade Cs as possible was clearly linked to accountability measures and their associated pressures, this was not always the case and some schools reported using entry strategies to help maximise their students' life chances. Thus, accountability at the school level was not the sole driver. Frequently, this desire to achieve results for the students ran alongside the desire to achieve results for the school:

> I mean the real drive was to get the best for the kid and the weaker kids, the borderline kids, we just said 'well if you've got the opportunity, give them as many chances as you can'. (Teacher, Academy A)

It is worth considering that, until recently, what has been in the best interests of the school and borderline students (in terms of achieving a threshold grade at least), have been closely aligned. Thus, using strategies to maximise achievement at grade C has been beneficial for both the school and the students: a win–win/low-risk situation:

> … the way I've viewed it in previous years is that in one respect it's great for the kids because they get lots of practice but for the school it's almost like they've banked C grades … it's a win/win situation …. (Teacher, Comprehensive A)

Following the changes to accountability announced in September 2013 (DfE, 2013b) only a student's first result will count in school performance indicators. Consequently, maximising the chances of achieving a grade C (through multiple entries) could be beneficial for the student, but not the school, requiring schools to balance the needs of the school and the students—a tension identified elsewhere (Ball et al., 2012). For those interviewed, the needs of the school appeared to take precedence. This resulted in the majority of schools withdrawing students who were due to enter the November 2013 examination series, due to the risk this was deemed to pose for results. Further, respondents suggested that, henceforth, they would enter students once at the end of Year 11, rather than using early and multiple entry:

Downloaded by [Ofqual] at 03:22 04 January 2018

> … we don't want them taking it if they're going to get a D, because we need Cs …. (Teacher, Academy H)

> I mean the whole of our Year 11 was entered in November and every single one was withdrawn. (Teacher, Academy A)

Only one respondent reported entering students in the November 2013 series as planned, due to the likely negative effects that withdrawal would have on their students. This was one of the higher-achieving schools and it seems reasonable to speculate that this school may have experienced less pressure to achieve results than their lower-achieving counterparts, potentially making their decision more straightforward:

> One of the reasons we chose not to withdraw is because of the impact it would have had on parents and students because they're all ready for it. They'd been prepared for it …. (Examinations Officer, Academy K)

The need to prioritise the schools' needs above the students' needs is not surprising, given the potential consequences of schools not achieving the required results, such as increased monitoring, inspections, and, ultimately, closure. Furthermore, the consequences for the school are likely to be wider-reaching: if the school were to close, there would be significant implications for the students as well:

> … I don't think this school or any other school doesn't care, I think they do. However, life is so much harder, you indirectly affect your students if you don't have a good … if your school isn't shown to be doing well. (Head of Department, Academy C)

Entry strategies—who, what, when and where?

The focus on results, primarily the grade C boundary, resulted in schools adopting a highly strategic approach to their entry practices for GCSE Mathematics. It was clear from discussions that, although different individuals within the schools were responsible for entry strategies and that these strategies varied significantly between and within schools, considerable time, thought, and effort was invested by each school in devising the strategy that they perceived would maximise results. This aligns with previous research that describes the 'interrogation' of school life designed to maximise results (see Ball et al., 2012; Gillborn & Youdell, 2000).

By far the most popular strategy reported here was early and multiple entry. All respondents, bar one, reported using this approach in the past for some or all of their students. The exception to this was the selective school that, likely under less pressure to achieve results due to the high calibre of their students, reported entering their students once at the end of Year 11. For the other schools, early entry was typically used to facilitate multiple entry across different examination series and examination boards, a practice that was likely to cease following the announcement that only a student's first entry attempt would count towards performance indicators (DfE, 2013b). The use of multiple examination boards was not unusual and there was evidence of some students having been entered with three UK awarding bodies within one academic year. This behaviour might be unique to Mathematics, since there is no internal assessment and the content across specifications is broadly comparable. It is doubtful whether this approach could be so easily replicated in other subjects:

Downloaded by [Ofqual] at 03:22 04 January 2018

> … they will have sat in November with AQA, they sat in January with OCR, they sat in March with Edexcel and then they sat with AQA in June so they will have had 4 entries in Year 11. (Teacher, Comprehensive B)

There was also evidence of students entering multiple specifications within the same examination series, either with the same, or different, awarding bodies. Again, this might be unique to Mathematics, since multiple specifications are available.[8] Whilst this approach increased the number of opportunities for students to enter, concerns were raised about the potential effects on the students involved:

> … we did have a couple, not very many, but we did have a couple who we gave a shot of the Unit 3 and the linear at the same time you know, that was a tough ask for those particular kids but … it just gives them another opportunity. (Head of Department, Academy F)

The frequent use of early and multiple entry was highly strategic and perceived as a means of allowing students, particularly C/D borderline students, as many entry attempts as possible through Years 10 and 11. Entering GCSE Mathematics four times or more was not uncommon:

> So, if I go back to last year, our Year 11s, the vast majority of the students were entered five times across the course of Year 10 and 11. So they were basically entered at every point they could have been right from the start of Year 10. (Head of Department, Academy D)

Within the practice of entering students early and on multiple occasions, schools were also highly strategic in choosing which examination to enter their students to. For example, some schools reported switching examination boards part-way through the course, or moving their lower-achieving students from a modular to a linear specification if they had performed poorly on the early modules and could not, or were unlikely, to achieve their target grade:

> I had pupils doing the modular course and if they weren't doing very well in their modules and it looked as though the modular route might lead to a lower grade then they've been entered or switched to enter for the linear model. (Teacher, Academy G)

Similarly, determining a student's tier of entry was often a strategic decision, although again, practices varied. Some schools opted to use the higher tier for their borderline grade C students, which they considered 'easier' for achieving a grade C due to the low number of marks required. This often led to weaker students being 'trained' specifically for the assessment:

> … if you've got a pupil on a lower level and they've been identified, they will be trained to do the higher paper because it is deemed as easier to get a C grade on than a foundation. (Teacher, Comprehensive A)

Other schools used tiering as a means of supporting progression, by entering their borderline students for the foundation tier initially then moving them to the higher tier to improve their grade. This was considered beneficial for the students, since without early entry, these students would have remained on foundation tier, capping their achievement at a grade C:

> … if we hadn't entered them at Year 10, we would have sat on them, they would have done Foundation at the end of Year 11 if we'd not done any early entry. (Teacher, Comprehensive B)

Tiering was also used strategically for more able students, typically with the aim of securing a grade C. Here, schools reported entering students to the foundation tier initially to 'bank' a grade C, before moving them to the higher tier. Although this approach mirrors the strategy discussed above, for the weaker students it was considered as a means of supporting

Downloaded by [Ofqual] at 03:22 04 January 2018

progression and benefiting students, yet for the more able students, it was done for the benefit of the school. Consequently, entering able students early at the foundation tier did not always sit well with respondents:

> As I say, the school that I trained at was very much, Year 9 some of their children entered Foundation, the top end entered Foundation in Year 9 to get a C which always seemed really bizarre to me because I was like they're the kids that you can pretty much bank will get a C or above no matter what …. (Teacher, Academy H)

The strategic manner in which tiering was used to support progression or bank grades is likely to be symptomatic of early and multiple entry. In a system of single examination entry, Gillborn and Youdell (2000) discussed the strategic use of tiering, yet here teachers reported 'playing it safe' and only entering students to a higher tier if they were likely to achieve well above the lowest grade available. Thus, early and multiple entry can facilitate movement between tiers in a way that is not possible with single examination entry.

Early entry as a standalone strategy

Although early entry was mostly used to allow students multiple entry attempts, there was some evidence of early entry being used as a strategy in itself. A small number of schools reported entering their highly able students early with the intention that they would sit another qualification the following year. This aimed to bridge the gap between GCSE and A-level for those students who intended to continue studying Mathematics in post-compulsory education. For the most able, this was often felt to be a successful strategy. However, achieving a high grade at GCSE when entering early was not always a given, even for the highly able, meaning that deciding who should, or should not, be accelerated in this manner, was not straightforward:

> With the top set kids … a lot of them got their A*s early so we were able to do a further maths qualification … which they really enjoyed. Some of them weren't able enough to do it so, again, you had kids in there who were getting put in for this exam who weren't probably clever enough for and bombing out and that and having sort of a negative view of maths from that experience. (Teacher, Comprehensive C)

Early entry was also used for what respondents described as 'at-risk' students—those students who were unlikely to remain in education until the end of Year 11. This practice aimed to ensure that these students achieved a grade before leaving school and was likely to continue despite the changes to accountability, with respondents considering this approach to be in the best interests of these particular students:

> … the other things that we'd use early entry for, and we would still use early entry for, is what we call 'kids at risk'. In other words, we don't think they're going to last until the end of Year 11. (Teacher, Academy A)

Flexible and experimental strategies

In addition to entry strategies being highly individualistic and complex, it was apparent that entry strategies were not static; rather, they changed and developed over time. Schools were often experimental in their approaches, attempting to devise what they perceived to be the most effective formula for achieving results, be it for the school, or for the students:

Downloaded by [Ofqual] at 03:22 04 January 2018

> Yes, we adapted … when we entered them, how, how we prepared them, what resources we put in place, all of that was on-going and they were … it was strategic decisions, it wasn't sort of like 'let's have a go with this scattergun on these and these'. It felt like that a bit at the start, it was a trial, it was an experiment, but then it became far more sort of set out and planned and delivered, yes. (Head of Department, Academy J)

This experimental approach often entailed trialling strategies with a particular group of students, before extending it to a larger cohort. In addition, strategies were constantly reviewed:

> … we opted to put a small cohort in for the linear maths in addition to them still following modular to see how effective it would be for certain types of students. We found that the results were quite promising and so last year we actually extended that to the whole cohort. (Examinations Officer, Academy K)

Entry strategies were developed in response to a number of factors, including the availability of assessments (which has frequently changed in GCSE Mathematics over recent years), the success or otherwise of previous strategies (as judged by examination outcomes; see Gillborn & Youdell, 2000), and external pressures such as accountability measures:

> … it's like 'right, well we'll try that' and then you do it once and you're slightly successful and then the … deputy head's like 'well we'll try it again, let's do it again next year'. (Teacher, Comprehensive A)

The experimental approach to entry strategies was often highly data-driven. This was facilitated by the use of multiple entry, since schools had data on each student's performance following each examination series which could feed into future teaching and entry strategies. The level of analysis undertaken was often striking; schools invested considerable time in tracking their students throughout the GCSE years and targeting interventions where necessary. To some extent, this offered a degree of personalised learning, since teachers could target the areas where individual students had performed poorly. The loss of such data following the move to a single examination entry was of great concern to respondents, since this was likely to result in increased workloads and fewer opportunities to support students in this manner:

> We used to do a lot of analysis of the results from the early entries and then we'd hit key areas that groups were weak on and we haven't got the opportunity to do that now so it means that we have to do that internally which is a massive admin task to mark it and uploading the results for every single pupil on a question by question basis. (Head of Department, Academy D)

Disquiet amongst teachers

The strategic manner in which students were entered for assessments was not without consequence and it was clear that the strategic practices used within schools did not sit well with many respondents. This has been alluded to in certain quotes above, yet it is worth making the point explicit, since it would be remiss to suggest that the use of strategic entry practices did not cause concern for respondents. Rather, it was clear that, although teachers were adopting highly strategic behaviour to maximise results, this was frequently coupled with an overwhelming sense of frustration and disquiet. This reflects previous findings concerning strategic entry to tiered assessments, where teachers felt forced to act in a way that was not necessarily in the best interests of the students (Gillborn & Youdell, 2000). In the

Downloaded by [Ofqual] at 03:22 04 January 2018

present study, particular frustration was expressed at the concept of training students to pass examinations and the high levels of support given to students:

> There was no emphasis on 'right, it's your exam, you've got to revise for it', none of that because they were so terrified that they wouldn't do it … we had to hold their hand the whole time and walk them through how to revise and spoon-feed them revision materials and I hated that aspect of it. (Teacher, Comprehensive C)

Many respondents reported experiencing a moral dilemma. On the one hand, respondents considered it their moral duty to get students a grade C, so would use whatever methods necessary to achieve this. However, on the other hand, individuals considered the use of certain entry and teaching practices—enacted for the benefit of the school—as immoral, due to the negative implications for students' learning. These two imperatives seemed to be pulling in opposing directions. This resonates with previous findings. Gillborn and Youdell (2000) discuss the dissonance between the morally/socially acceptable approach and the approach that will produce the best results when discussing setting students by ability (see also Perryman et al., 2011):

> I mean it was an issue that was a real bugbear for me when I was at my last school. I didn't really agree with this sort of racing to mediocrity. (Teacher, Comprehensive C)

This tension may have been particularly acute for the respondents in these interviews, since they were largely classroom teachers who were interacting with students on a regular basis. Had the views of senior management been gathered, the responses may have been different:

> I mean I don't think any of it was good. I didn't enjoy it. I didn't want to do it and I'm glad I'm not having to but I don't know if I was a Head of one of those schools that I would've necessarily done anything any different. (Head of Department, Academy J)

Responsibility

The final theme relates to the issue of responsibility and where responsibility for the situation described here lies. This issue was not included in the interview schedule, but arose frequently in the data. For many respondents, responsibility was placed firmly on the Government and there was a sense that schools were merely acting as agents within the assessment system to something that was being dictated to them. Interviewees felt that their hands had been forced by the constant pressure in the system. Indeed, many individuals openly admitted to 'gaming the system'. The willingness to admit this so freely may be due to a sense that schools were not responsible for their actions:

> … we've been 'playing the game' for the Government and, actually, our pupils don't have a good knowledge of maths, they have a good knowledge of an exam …. (Teacher, Comprehensive A)

> … you know, it's not us … our school didn't make the policy on what was the right grade. If there wasn't a league table or it wasn't published then it might even be a better system but that's not the case so we have to push the Cs as our main driver and we do. (Head of Department, Academy C)

Many respondents reported that they were doing whatever they could to maximise results within the confines of the examination system:

> … we've really just responded to whatever the Government have put available and we, like every other school, will just take the best, the optimum route, through what the system allows. (Teacher, Academy B)

Downloaded by [Ofqual] at 03:22 04 January 2018

> … everyone was playing the numbers game so, last year, when we entered pupils 5 times, we did that because we could. (Head of Department, Academy D)

For some, there was an inevitably about this and it was felt that, once a measure of success was in operation, this became the goal to achieve:

> Well if you're going to be measured, if somebody says to you 'you're going to be judged on this', everything else goes out the window and you do what you have to, to be judged on whatever that is and that's whatever isn't it? (Head of Department, Academy J)

The sense that schools were merely reacting to Government policy was reinforced by the language that respondents used during the interviews, particularly regarding the accountability changes announced in September 2013. Many participants referred to these changes as 'banning' the use of certain entry strategies, when schools would still be free to use whatever strategy they wished. However, only a student's first attempt would count for performance indicators, meaning there were likely implications for the results that a school would achieve:

> What they've done is there's no flexibility at all; they've really just said no early entry. (Examinations Officer, Academy K)

Concluding comments

The findings from this study reinforce how the pressure of performance measures have been driving behaviour in schools, prompting the use of early and multiple examination entry to maximise results. This has been coupled with considerable disquiet amongst those implementing these strategies: classroom teachers and heads of departments. The study also provides evidence of the way in which strategies such as early and multiple entry have been used for other purposes: to accelerate the most able students to further study, or to ensure that 'at-risk' students (who are unlikely to remain in education) gain a qualification. Since the study was largely exploratory and gathered the views of only a small number of individuals, no claims concerning the generalisability of the findings are made. However, anecdotal evidence and media coverage suggest that the strategies identified here are unlikely to be unique to the schools that participated in this research (e.g. Paton, 2013).

The behaviour discussed in this paper centred upon the strategic entry of students to GCSE Mathematics assessments to maximise results, either for the purposes of schools or students. Whilst this paper focuses on early and multiple entry, the findings resonate with much previous research exploring the ways in which performance measures and their associated pressures can drive behaviour within schools (e.g. Ball et al., 2012; Gillborn & Youdell, 2000; Perryman et al., 2011). This behaviour has been described as 'gaming' (e.g. Goldstein, 2008; Wolf, 2011) and is not unique to the field of education (e.g. see Burgess, Propper, & Wilson, 2005), or the English assessment context (e.g. see Hutchings, 2015, for a summary of findings from the US).

By focusing on early and multiple examination entry, this paper also adds to the limited academic literature exploring these behaviours. Whilst the use of these strategies has declined following recent policy changes (DfE, 2013b; Ofqual, 2014), their existence has likely had a number of consequences for teachers, schools, students, and the wider examination system, yet little research has focused on this. Many of these consequences are likely to transcend the behaviours discussed here and be symptomatic of strategic

Downloaded by [Ofqual] at 03:22 04 January 2018

behaviours adopted more widely within schools, highlighting the need to consider such consequences when exploring schools' reactions to accountability measures. Furthermore, including additional perspectives—such as the student perspective—is also likely to be useful (e.g. see Gillborn & Youdell, 2000). This evidence is often lacking (Elwood, 2012), but offers an important avenue for further study. In particular, considering the effect of strategic entry on different types of students is important, since the effects are likely to be both positive and negative, raising issues around equity and fairness (e.g. see Gillborn & Youdell, 2000).

At the time of writing this paper, significant changes were being made to the English assessment system, including the introduction of new GCSEs for first teaching in 2015, and the introduction of new accountability measures from 2016 (Laws, 2013). Thus, there is scope for these policy changes to drive other behaviours in the future (Stobart, 2008), with similar drivers and consequences. It is therefore hoped that the current study, alongside the existing literature, can be informative, and evoke caution, during future periods of reform.

Downloaded by [Ofqual] at 03:22 04 January 2018

Notes

1. Note that from 2016, a new measure, Progress 8, will be used for accountability purposes (Laws, 2013). This measures students' progress across eight GCSE subjects.
2. GCSE Mathematics examinations are tiered; the foundation tier allows access to the lower grades (C–G), and the higher tier, the higher grades (A*–G).
3. Throughout this paper early and multiple entry refers to GCSE certification, not entry to modules within a qualification.
4. Ofsted, the Office for Standards in Education, Children's Services and Skills, are responsible for school inspections and monitoring.
5. In England there are three examination boards offering GCSE qualifications—AQA, EdExcel, and OCR.
6. Academies are publicly-funded, independently-run schools. Comprehensives are state schools that do not select their intake based on attainment.
7. At the time the research was conducted, GCSE Mathematics was available in January, March, June, and November examination series.
8. Until November 2013, linear, modular, and linked pair pilot specifications were available.

Acknowledgements

The author would like to thank Kate Kelly and Victoria Spalding for their assistance with the data collection and analyses.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributor

Rachel Charlotte Taylor is a senior research associate at the Centre for Education Research and Practice at AQA and a DPhil student at the Oxford University Department of Education. Her research focuses on educational assessment. Current research interests include the influence of policy change on entry and teaching strategies for GCSE assessments and the implications of this for maintaining standards.

References

Advisory Committee on Mathematics Education. (2011). *Early and multiple entry to GCSE Mathematics*. Retrieved from http://www.acme-uk.org/news/news-items-repository/2011/5/position-paper-on-early-and-mutiple-entry-to-gcse-mathematics

Ball, S., Maguire, M., Braun, A., Perryman, J., & Hoskins, K. (2012). Assessment technologies in schools: 'Deliverology' and the 'play of dominations'. *Research Papers in Education, 27*, 513–533.

Bell, J. F. (2001). Patterns of subject uptake and examination entry 1984–1997. *Educational Studies, 27*, 201–219.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*, 77–101.

Bryan, K., & Digby, A. (1983). Pupil performance and school effectiveness. *Westminster Studies in Education, 6*, 33–41.

Burgess, S., Propper, C., Slater, H., & Wilson, D. (2005). *Who wins and who loses from school accountability? The distribution of educational gains in English secondary schools*. Bristol: The Centre for Market and Public Organisation.

Burgess, S., Propper, C., & Wilson, D. J. (2005). *Will more choice improve outcomes in education and health care? The evidence from economic research*. Bristol: The Centre for Market and Public Organisation.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). Oxon: RoutledgeFalmer.

de Wolf, I. F., & Janssens, J. G. (2007). Effects and side effects of inspections and accountability in education: An overview of empirical studies. *Oxford Review of Education, 33*, 379–396.

Department for Education. (2010). *The importance of teaching: The schools white paper*. London: DfE.

Department for Education. (2011). *Early entry to GCSE examinations*. London: DfE.

Department for Education. (2013a). *Expected levels of school and college performance*. Retrieved from http://www.education.gov.uk/schools/performance/fs_13/index.html

Department for Education. (2013b). *Changes to early entry at GCSE*. Retrieved from https://www.gov.uk/government/news/changes-to-early-entry-at-gcse

Department for Education and Skills. (2001). *Schools achieving success*. London: DfES.

Department for Education and Skills. (2005). *14–19 Education and skills*. London: DfES.

Department of Education and Science. (1988). *Education reform act*. London: HMSO.

Elwood, J. (2012). Qualifications, examinations and assessment: Views and perspectives of students in the 14–19 phase on policy and practice. *Cambridge Journal of Education, 42*, 497–512.

Gill, T. (2010). *How old are GCSE candidates?* Statistics Report Series No. 20. Cambridge: Cambridge Assessment.

Gillborn, D., & Youdell, D. (2000). *Rationing education: Policy, practice, reform and equity*. Buckingham: Open University Press.

Goldstein, H. (2004). Education for all: The globalization of learning targets. *Comparative Education, 40*, 7–14.

Goldstein, H. (2008). Evidence and education policy—some reflections and allegations. *Cambridge Journal of Education, 38*, 393–400.

Hutchings, M. (2015). *Exam factories? The impact of accountability measures on children and young people*. London: National Union of Teachers.

Jacob, B. A., & Levitt, S. D. (2003). *Rotten apples: An investigation of the prevalence and predictors of teacher cheating*. National Bureau of Economic Research.

Laws, D. (2013). *Reforming the accountability system for secondary schools*. Retrieved from https://www.gov.uk/government/speeches/reforming-the-accountability-system-for-secondary-schools

Mansell, W. (2007). *Schools risk becoming 'exam factories'*. Retrieved from http://www.telegraph.co.uk/news/uknews/1561167/Schools-risk-becoming-exam-factories.html

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach*. London: Sage.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. London: Sage.

Noyes, A., Drake, P., Wake, G., & Murphy, R. (2010). *Evaluating mathematics pathways: Final report*. London: DfE.

Downloaded by [Ofqual] at 03:22 04 January 2018

Noyes, A., & Sealey, P. (2011). Managing learning trajectories: The case of 14–19 mathematics. *Educational Review, 63*, 179–193.
Ofqual. (2014). *Summer GCSE and IGCSE entries for England: Provisional figures April 2014*. Coventry: Ofqual.
Ofsted. (2013). *Schools' use of early entry to GCSE examinations: Its usage and impact*. Manchester: Ofsted.
Paton, G. (2013). *School standards threatened by 'early and often' GCSEs*. Retrieved from http://www.telegraph.co.uk/education/educationnews/10250548/School-standards-threatened-by-early-and-often-GCSEs.html
Perryman, J., Ball, S., Maguire, M., & Braun, A. (2011). Life in the pressure cooker—school league tables and English and mathematics teachers' responses to accountability in a results driven era. *British Journal of Educational Studies, 59*, 179–195.
Pope, S., & Noyes, A. (2011). Early entry in GCSE Mathematics. *Proceedings of the British Society for Research into Learning Mathematics, 31*, 127–131.
QCA. (2009). *Changes to GCSE Mathematics*. London: QCA.
Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics, 92*, 1394–1415.
Stobart, G. (2008). *Testing times. The uses and abuses of assessment*. Abingdon: Routledge.
Strand, S. (2012). The White-Black Caribbean achievement gap: Tests, tiers and teacher expectations. *British Educational Research Journal, 38*, 75–101.
The Mathematical Association. (2010). *On the subject of early and repeated presentation for GCSE Mathematics*. Retrieved from http://www.mei.org.uk/index.php?page=news&newsid=236
The National Strategies. (2009). *Secondary Mathematics update: Annex 2 early entry to GCSE Mathematics*. London: DfE.
West, A. (2010). High stakes testing, accountability, incentives and consequences in English schools. *Policy and Politics, 38*, 23–39.
West, A., Mattei, P., & Roberts, J. (2011). Accountability and sanctions in English schools. *British Journal of Educational Studies, 59*, 41–62.
West, A., & Pennell, H. (2000). Publishing school examination results in England: Incentives and consequences. *Educational Studies, 26*, 423–436.
Wilson, D., Croxson, B., & Atkinson, A. (2006). What gets measured gets done: Headteachers' responses to the English secondary school performance management system. *Policy Studies, 27*, 153–171.
Wilson, D., & Piebalga, A. (2008). Performance measures, ranking and parental choice: An analysis of the English school league tables. *International Public Management Journal, 11*, 344–366.
Wolf, A. (2011). *Review of vocational education: The Wolf Report*. London: DfE.

Downloaded by [Ofqual] at 03:22 04 January 2018

Appendix G – interview schedule

THEME	PROMPTS
Establish common understanding of early/multiple entry	Clarify meaning of early and multiple entry. I.e. early entry refers to certification before the end of Year 11 (not early entry to modules) and multiple entry could refer to repeated entry (e.g. following early entry) as well as entering for multiple specs/with multiple awarding bodies in the same series.
Description of entry strategy until June 2013	Recently there have been various changes to accountability measures. Could you describe the entry strategy used in your school for GCSE mathematics prior to this (i.e. until June 2013), with reference to the use of early and/or multiple entry? • *Were entry strategies devised at the whole cohort or individual level?* • *Did entry strategies differ depending upon the ability of the students/tier of entry/the grade that students are expected to achieve?* • *Did your entry strategy include multiple entry between exam boards? Use of modular/ linear specs/switching between?* • *What re-sit opportunities were there for early entry students?* • *What provision was there for students following early entry who didn't re-sit?* • *How did post-early entry provision differ depending on the grade that a student achieved? Particularly those above/below the grade C threshold?* • *Did your strategy develop over time? In response to what?*

Impact of changes to accountability on entry strategies	The DoE have recently announced changes to accountability measures (i.e. that only the first entry attempt for students will be included in performance measures and the revision of the info included in performance tables). How do you think this will influence your entry strategy? • *Will anything change?* • *Will you still use early/multiple entry for students? If so, who?* • *What about multiple entry between exam boards in the same series?* • *Will any changes influence how students are taught and prepared for assessments?* • *What about re-sitting – will this be used in the same way? E.g. if a student gets a D at the first attempt, how motivated do you think schools will be to re-enter them?* • *How do you perceive this change influencing the results for your centre?* • *This change has been introduced alongside rising floor targets– how do you feel about this?* These changes are intended to discourage inappropriate use of early entry. Do you support early entry being discouraged in this way? • *Should it be discouraged?* • *Do you think there should be rules governing the use of early entry? If so, what else could they be?* • *Are there other/better ways to discourage early/multiple entry?* The changes are intended to reduce behaviour that has been described as 'gaming' of the system. Do you think they could drive other behaviours? • *What could be the unintended consequences of these changes?* • *Could this impact negatively on students at all?* •
Process for devising entry strategies	Could you talk me through the process in your school for devising the entry strategies that you've just described? • *Which individuals are involved in this process? At what level are decisions made?* • *Are there pressures from senior management/the LEA (<u>not</u> Academies though)?* • *Do the students/parents have any input?* • *Do you review/change strategy? On what basis?*
What factors influence entry strategies and the impact of policy changes	What have been the main drivers influencing your entry strategies for GCSE mathematics? What did you hope to achieve by using your particular entry strategy? • *Was early entry used for acceleration/enrichment for the most able students? What about weaker students?* • *Are there pressures from senior management/LEA?* • *Do the students/parents have any input?*

	• *What do you perceive to be the positive outcomes of early/multiple entry? Did you have any concerns about using this approach? Did you perceive any risks? To the individual/system?* • *Do you think your entry strategy has impact upon the results in your centre? If so, how?* • *Educationally, do you think there are any risks to the education system (as a whole) due to the use of early/multiple entry?* • *What influence do you think these strategies might have on results at the national level?* There have been a number of policy decisions over recent years that are thought to have influenced entry strategies (e.g. removal of KS3 tests, the introduction of modularity at GCSE, the introduction of the EBacc and changes to accountability measures) – do you think these changes have influenced schools' entry strategies? If so, how? • *What about your own entry strategy?* • *Which policy in particular do you think has been particularly influential?* How effective do you consider your entry strategies to have been? • *In terms of outcomes and how much students learn? And in terms of preparing students for future study/progression?* • *Is the success of entry strategies actively reviewed?* • *Were strategies adapted over time? On what basis?*
Influence of entry strategies on teaching strategies	Could you describe any impact that your entry strategies have on the way in which students are taught/prepared for assessments? • *Is the same amount of time given to KS4 material for early entrants as for those entering at the end of Year 11?* • *Are certain topics/techniques prioritised?* • *Do they start the KS4 programme of study early?* • *Accusations of teaching to the test when entering students early – perceptions of this?* • *Does this differ to students entering at the end of Year 11? If so, how?* • *What do you think this means for what students know and can do?*

Appendix H – transcripts and code book

This table provides further information about the transcripts that were coded including reference to the school, the participants, the date the interview was conducted, and the word count of each transcript.

School	Date	Participants	Word count
Comprehensive A	18/11/2013	Examinations Officer and GCSE mathematics teacher	8,751
Academy A	26/11/2013	GCSE mathematics teacher	6,169
Academy B	27/11/2013	GCSE mathematics teacher	4,161
Academy C	29/11/2013	Head of department	6,939
Academy D	04/12/2013	Head of department	7,962
Academy E	04/12/2013	GCSE mathematics teacher	8,598
Academy F	05/12/2013	Head of department	6,097
Comprehensive B	10/12/2013	GCSE mathematics teacher	7,576
Academy G	08/01/2014	GCSE mathematics teacher	5,311
Academy H	08/01/2014	GCSE mathematics teacher	5,388
Selective A	16/01/2014	GCSE mathematics teacher	5,297
Comprehensive C	16/01/2014	GCSE mathematics teacher	7,265
Academy J	16/01/2014	Head of department	5,225
Academy K	17/01/2014	Examinations officer	11,388

Table H1. *Dates, participants and word counts for the interviews.*

This log shows the codes that emerged from the data, linked to the over-arching framework that was developed to organise the themes (see Chapter 5, Figure 13). The code book also provides an explanation of the code and an example from one of the interviews. Note that due to the complexity of the data some of the codes related to multiple themes. Furthermore, the final two themes – disquiet and responsibility – are presented outside of the framework so the codes are mapped to these themes instead.

Table H2. *Coding log and mapping to over-arching framework*

Over-arching framework	Code	Sources	References	Explanation	Example
Drivers	Results, performance measures and Government.	14	65	Influence of Government and performance measures and associated pressure on schools' results	'Definitely the key drivers, accountability and the need to meet targets'
	Grade C	13	72	Grade C and the importance of this threshold	'… obviously to achieve that grade C is very important if they've got the ability to do it because it opens doors of opportunity
	Progress	11	36	Students making progress for levels of progress measures or to improve their grade	'Now there's greater awareness of measurement by pupil progress rather than just hitting the grade C mark' 'Levels of progress is the big one now'
	Prioritising students	9	26	Desire to provide students with opportunities and support to ensure they achieve a particular grade	'As far as I'm concerned, it's for the kids'
	Parents	9	15	Levels of parental input	'I had one parent contact about entries for the entire cohort of Year 11'
Approaches	Early entry, multiple entry and EMEE (coded together due to extensive overlap)	14	74	Entering students early or multiple times before the end of Year 11, and early entry facilitated by multiple entry	'It's been nearly the entire Year 10 would do it in the summer' 'So they were basically entered at every point they could have been right from the start of Year 10 and 11'
	Multiple exam boards or specifications	13	18	Entering multiple specifications or exam boards, either in the same or different examination series	'… certainly taken the opportunity to … not very frequently but in the past I have used that opportunity of trying different exam boards'

	Implications of accountability changes	12	50	Implications of changes to accountability measures announced in September 2013 for entry and teaching approaches	'… originally, in November, everyone who didn't hit either the A or A* or the C were going to retake it in November. However, that all changed when we came back …'
	Decision making and decision makers	12	17	How decisions are made within schools about entry approaches and by whom	'Well I think in this school it tends to be made mainly by the senior teachers who are responsible for data collection and response for standards'
	Exam preparation and training	11	34	Exam training, revision cycles and targeted interventions	'… throughout Year 11, became an exercise of just revision constantly because you were just preparing for the next exam that was only a couple of months away each time'
	C/D borderline students	10	26	Focus on students on the C/D borderline	'… so really, the big repeated entry was those C/D borderline kids'
	Teaching groups	9	26	Approaches to arranging students into teaching groups and moving students between groups	'So what that did was allow us to move a lot of pupils around between groups so, when pupils hit target grades of C's, they could move to B focus groups and so on'
	At-risk students	9	16	Students not considered likely to remain in education until the end of KS4	'And the other things that we'd use early entry for, and we would still use early entry for, is what we call 'kids at risk'. In other words, we don't think they're going to last until the end of Year 11'
	Studying other qualifications or subjects	9	20	Studying other mathematics qualifications or focusing attention on other subjects following EMEE	'… if they ended up getting a grade C in the January exam … if they wanted to then they could stop doing maths and they were able to use that maths time for revision and work on other subjects' 'So that's something we're quite passionate about is making sure that they carry on doing maths in some form …'

	Data and tracking students	8	16	Using data and tracking students to target interventions and feedback following entry attempts	'… we are obviously very data strong and what we're doing is analysing our data'
	Flexible entry approaches	7	12	Flexibility in adapting entry approaches	
	Higher ability students	7	11	Provision and entry approaches for higher ability students, including early entry and enrichment	'One strategy was enhancement materials to prepare them for doing maths in the sixth form'
	Re-sitting to improve grades	6	7	Re-sitting to improve grades following an early entry	
	Topic spotting and predictability	6	16	Targeting teaching to certain topics where marks were likely to be scored or topics likely to come up on the assessments	'… they are prioritising the topics that a) we know definitely come up because they come up every year …'
	Modular and linear entry	6	6	Entering both the linear and modular specifications or switching partway through the course	'I had pupils doing the modular course and if they weren't doing well in their modules … they've been entered or switched to enter the linear model'
	Foundation then higher entry	6	13	Entering students to foundation tier then moving to higher tier, including to 'bank' a grade C or to support progression	'… I think the policy of putting pupils that are at Higher level in for foundation to make sure that a grade C was banked … my judgement on that would be probably a bit of a waste of money'
	Higher tier entry for borderline C/D students	5	9	Entering borderline C/D students to the higher tier	'If you think the kid's got an inkling of a chance of getting a C, you put them in for the higher'
	Higher tier is easier	3	5	Perception that it is easier to gain a grade C on the higher tier	'… you put them in for the Higher level because they've got less marks to score'
Implications	Perceptions of accountability changes	13	40	Perceptions of changes to accountability measures announced in September 2013, e.g. support and concerns	'So I do support the removal of early entry because I do feel it's probably been abused in certain schools' 'So I think in some respects the changes that are being made are unfair on students where early and multiple entry suits them'

	Student motivation	11	44	The impact of entry and teaching approaches on students' motivation and engagement	'… some of the people were quite motivated as well when they saw how close to a C they'd been… ' 'At the moment we've got a big motivation issues with Year 11, particularly the ones that were in the group where some have gone because they got As and A*s'
	Learning	10	23	Implications of EMEE for student learning	'Our students don't have a good knowledge of maths'
	Outcomes	10	19	Implications of EMEE for student and national examination outcomes	'… So I think it must have positively influenced C and above certainly'
	Potential	9	14	Students not realising their potential through EMEE, particularly the middle achievers	'I've seen so many students who are so good mathematically … but just not going to make any progress because all they needed to do … was to get a grade C …'
	Maturity	5	6	Implications of EMEE for younger students who might be lacking maturity	'What you tended to find was there was sometimes pupils who weren't socially mature enough to realise 'this is the real GCSE' …'
	Cost	4	4	Budget and cost implications of EMEE	'… the amount of money we were spending on exam entries probably wasn't ideal'
Disquiet	Disquiet	10	26	Disquiet about entry and teaching approaches used in schools	'It's not really educating them, it's training them to sit an exam so that didn't sit well with me …'
	Change	6	10	Changes and reforms to the assessment system and the implications of this	'There's more reform and there's been that much change in maths'
Responsibility	Gaming	9	11	Gaming within the confines of the exam system	'At the end of the day we're all playing a big game'

Appendix J – simulation output

Table J1: *Differences between best grades and all grades according to the proportion of multiple entrants*

Proportion of multiple entrants	*	A	B	C	D	E	F	G
0.05	0.11	0.22	0.29	0.30	0.25	0.16	0.09	0.03
0.10	0.22	0.43	0.59	0.61	0.49	0.33	0.17	0.06
0.15	0.33	0.65	0.89	0.90	0.73	0.49	0.26	0.09
0.20	0.44	0.87	1.18	1.21	0.98	0.66	0.34	0.12
0.25	0.56	1.09	1.48	1.51	1.23	0.82	0.43	0.15
0.30	0.67	1.30	1.77	1.81	1.47	0.98	0.52	0.19
0.35	0.78	1.52	2.07	2.11	1.72	1.15	0.60	0.21
0.40	0.90	1.75	2.37	2.42	1.97	1.32	0.70	0.25
0.45	1.01	1.97	2.67	2.72	2.21	1.48	0.78	0.27
0.50	1.11	2.17	2.96	3.02	2.45	1.64	0.86	0.31
0.55	1.23	2.40	3.26	3.32	2.70	1.80	0.94	0.34
0.60	1.34	2.62	3.55	3.63	2.95	1.97	1.04	0.37
0.65	1.45	2.83	3.85	3.93	3.19	2.13	1.12	0.40
0.70	1.56	3.05	4.14	4.24	3.43	2.29	1.21	0.43
0.75	1.67	3.27	4.44	4.54	3.68	2.47	1.30	0.46
0.80	1.79	3.49	4.73	4.84	3.93	2.63	1.39	0.49
0.85	1.90	3.71	5.03	5.13	4.19	2.79	1.47	0.53
0.90	2.02	3.92	5.33	5.44	4.42	2.98	1.57	0.56
0.95	2.14	4.14	5.62	5.75	4.67	3.14	1.66	0.60