

# The SES equity gap and the reform from modular to linear GCSE mathematics

**Anne Pinot de Moira\*** (Independent Consultant), **Michelle Meadows** (Ofqual) and **Jo-Anne Baird** (University of Oxford)

**\*corresponding author**

Anne Pinot de Moira  
Independent Consultant  
Norham Fellow  
Department of Education  
University of Oxford  
15 Norham Gardens  
Oxford OX2 6PY  
[anne.pinotdemoira@gmail.com](mailto:anne.pinotdemoira@gmail.com)

Dr Michelle Meadows  
Executive Director Strategy, Risk and Research, Deputy Chief Regulator, Ofqual  
[Michelle.Meadows@ofqual.gov.uk](mailto:Michelle.Meadows@ofqual.gov.uk)

Professor Jo-Anne Baird  
Director, Department of Education, University of Oxford  
[jo-anne.baird@education.ox.ac.uk](mailto:jo-anne.baird@education.ox.ac.uk)

## **Author biographical statements**

**Anne Pinot de Moira** is a chartered statistician with over 20 years' experience working in the fields of assessment and education. She was Head of Assessment Research at the UK's largest awarding organisation, AQA, and is now an independent consultant. Anne is an Honorary Norham Fellow at the University of Oxford Department of Education.

**Dr Michelle Meadows** is Ofqual's Deputy Chief Regulator and Executive Director for Strategy, Risk and Research. She is responsible for research that supports the development of high-stakes assessment in general and vocational qualifications; successful delivery of reliable and valid public examinations; qualification design that stimulates high-quality teaching and learning; and setting and maintaining examination standards. She is also responsible for the evidence-based regulation of standards over time and between awarding organisations in England. Michelle has a PhD in psychology from the University of Manchester and is an Honorary Research Fellow at the University of Oxford Department of Education.

**Professor Jo-Anne Baird** is Director of the Department of Education at the University of Oxford. She previously held academic posts at the Institute of Education, University of London and the University of Bristol. Prior to this, she was Head of Research at the Assessment and Qualifications Alliance, where she managed the research programme and was responsible for the standard-setting systems for public examinations. Her first degree and doctorate were in psychology and she has an MBA. Her research interests are in educational assessment, including system-wide structures and processes, examination standards, marking and assessment design.

# The SES equity gap and the reform from modular to linear GCSE mathematics

## Abstract

This paper addresses whether introduction of end of course, linear General Certificate of Secondary Education (GCSE) examinations changed the socio-economic equity gap in England. The GCSE is a national examination offered in a wide range of subjects and taken by almost the entire 16-year-old age cohort. Between the years 2007 and 2014, it underwent a number of reforms to both the underlying curriculum and the examination structure. At the beginning of the period, examinations were primarily modular in structure where the course was decomposed into discrete units tested in a staged manner. By 2014, all GCSE examinations were linear and the whole course content was tested simultaneously when study was complete. These structural changes and the curriculum reforms mean that the impact of modular and linear testing on the performance of students has been the focus of recent interest. Some educational commentators suggested that modular examinations are more suitable for lower performing students, including those with lower socio-economic status (SES). This research has been conducted to monitor the socio-economic equity gap in the light of the structural changes. It focusses on GCSE mathematics and concludes that, although there is still a clear gap in attainment between disadvantaged students and their peers, this gap does not seem to be exacerbated by examination structure. In other words, the linearisation of GCSE mathematics is unlikely to have increased inequity between students of high and low SES.

## Keywords

modular assessment; linear assessment; GCSE; equity gap.

## Introduction

This article is based upon a project investigating the impact of reform of examination structures in England (Baird et al., 2019). Here, we focus specifically on performance in mathematics and on socio-economic attainment gaps in relation to two different examination structures. The examinations investigated are the General Certificate of Secondary Education (GCSE). GCSEs are normally taken by 16-year-olds following a two-year course of study in England, Wales or Northern Ireland. In this research, we consider only students who were 16-years-old and who were educated in state schools. These students represent ‘typical’ entrants and account for about three quarters of the entry cohort.

Students normally sit between eight and ten GCSEs (Gill & Williamson, 2016) and schools can choose which syllabuses to offer from 53 GCSE subject areas.<sup>i</sup> When GCSEs were introduced in 1988, they were the school leaving examination and catered for the entire ability range, unlike their predecessors (Certificate of Secondary Education and Ordinary

Levels). Since their introduction, they have undergone various reforms and testing has taken the form of both linear and modular examinations. In linear examinations, the course content is tested together at the end of the course of study, whilst with modular examinations the course is decomposed into discrete units that are taken in a staged manner. Modular assessment is prevalent in higher education and in vocational education but is not a common structure in secondary school examinations internationally. Of 18 jurisdictions investigated by Baird et al. (2019), three had modular examinations in secondary school (France, Queensland and Scotland) but one of those (Scotland) has since revised most of those examinations to a linear structure.

Of course, even linear examinations often involve more than one component, such as multiple question papers, and many GCSEs include practical assessments or coursework. There are different forms of modular examinations. Ertl and Hayward (2010) distinguished three forms of modularity: internal, external and connective. With reference to their classification, the modular GCSEs were mainly *internal* as the curriculum was decomposed into separate units that jointly formed the qualification. Typically, the units themselves were not available to other qualifications (an *external* definition) but there was a brief period in which some GCSE units were eligible as part of separate Diploma qualifications. A fully connective model was never used for GCSEs, as this would have undermined curricular coherence. Connective models permit accumulation of credit across a personalised curriculum.

Most GCSEs were linear in structure until 2009. In 2008, the examinations regulator published GCSE criteria that were permissive with respect to modularisation (Qualifications and Curriculum Authority, 2008). This marked a widespread move to modular GCSE examinations and, by 2010, almost all of the available GCSEs, from all five examination boards operating in England (AQA, CCEA, OCR, Pearson, WJEC), were modular in structure. However, mathematics GCSEs followed a different pattern where modular and linear syllabuses were offered both before and after the 2008 change to the GCSE criteria.

In 2012, the Coalition Government signalled a range of changes to examination policy. As a result of this, Ofqual, the regulator, required that all examinations for the existing GCSEs (modular or linear) had to be sat at the end of the course in a linear fashion from 2014. Some schools had entered pupils for the examinations in this manner in any case. At the same time, the (then) Minister for Education, Michael Gove, initiated a wholesale reform of GCSEs to be first sat in 2017. The new GCSEs were designed to increase the demands of examinations, which he considered to have been 'dumbed down', to have suffered from grade inflation, and to have resulted in a poor educational experience for young people (Hansard, 17 September 2012 col 655; Hansard, 7 February 2013 col 441). These reforms included the introduction of revised GCSE syllabuses and examinations, a key feature being higher standards of curriculum and assessment. Other changes included a reduction in coursework (known as controlled assessment), the removal of tiering in a number of subjects, and a move from letter grades to numerical grading<sup>ii</sup>.

The GCSE was formerly the school leaving age examination until near the end of the period under investigation. Until 2013, the school leaving age in England was 16, but since 2015 young people have to remain in education or training until age 18. In 2013 and 2014, young people had to be in education and training until age 17. In practice, however, approximately

95% of 16-17-year-olds and 90% of 16-18-year-olds stayed in education or training beyond the statutory school leaving age during this period (Department for Education, 2010, 2013, 2017).

Nonetheless, the GCSE is an important examination for young people, as it is used to select for access to further education courses and to some prestigious higher education institutions (e.g. Oxford and Cambridge universities) where they are considered alongside A-level results, typically taken at age 18. Additionally, GCSE results are important to schools as they are used by the inspectorate and in school performance tables (sometimes referred to as league tables). GCSEs are, therefore, high-stakes examinations for pupils, teachers and schools. School performance table targets have focussed on GCSE English and mathematics outcomes since 2006 (Leckie & Goldstein, 2016) and performance in these subjects is also included in the current English Baccalaureate<sup>iii</sup> and Progress 8 targets.<sup>iv</sup>

This research has been conducted to monitor the effect of the return to linear examinations after the reforms introduced from 2014. In particular, it considers the socio-economic equity gap with reference to the changes. It builds on the quantitative strand of a report considering examination reforms and the impact of linear and modular examinations (Baird, et al., 2018).

Comparisons of examination outcomes are analysed by the examination boards themselves, in both collaborative comparability studies and in the annual inter-board statistical screening programme. A small number of the comparability studies included modular and linear examinations, but none of them investigated GCSE examinations (D'Arcy, 1997; Qualifications and Curriculum Authority, 2006). The general finding from these studies was of no difference between modular and linear examinations, though one study in A-level English literature indicated that the modular examinations might have been more demanding at grade E, but not at grade A (D'Arcy, 1997). Whenever comparisons of examination outcomes give rise to concerns about standards, appropriate adjustments are made in subsequent years.

Before turning to the current study, we briefly outline empirical research on the effects of re-sitting, because this feature of the examinations is inextricably linked with the (modular or linear) structure and is often considered to be problematical in terms of maintaining standards. Since the current research looks at socio-economic status (SES) equity gaps, we discuss what was previously known about SES gaps in GCSE mathematics and GCSEs in general. To provide context, we explain the standard setting procedures applied to the current data because standard setting can have differential impacts on examinations, including modular and linear.

### ***Re-sitting effects in modular GCSE mathematics outcomes***

Although re-sitting is possible in linear examinations, all of the question papers must be re-taken in the same examination series. The frequency of re-sitting was controlled in modular GCSEs; for example, from 2004, pupils were only allowed to re-sit each module once (Qualifications and Curriculum Authority, 2004). Later, a new requirement was added so that 40% of the assessment had to be taken at the end of the course (Qualifications and Curriculum Authority, 2008).

Much of the research on modular examinations is on re-sitting in A-levels. These examinations were modular qualifications beginning with Curriculum 2000 (initiated by Dearing, 1996). Students did re-sit modules, but the uptake was far lower than the available offer, and re-sitting affected the resulting grades to a lesser extent than might have been predicted (e.g. Qualifications and Curriculum Authority, 2007).

Similar findings were reported by Vidal Rodeiro and Nádas (2012) for GCSE mathematics. They analysed results from one examination board, finding that students who re-sat units typically improved their marks on the second sitting. Such improvements could be a result of increased knowledge and skills, standard setting being more lenient, or the error associated with (all) measurements. As modules are taken in a staged manner, pupils mature between assessments, experience more of the curriculum and may have revised specifically for the curriculum content being assessed.

Previously, GCSEs often involved two question papers, giving approximately 20 examinations to be taken at the end of the course for a student taking 10 GCSEs. With the blanket modularisation, there tended to be more question papers and many more examination opportunities. A typical modular GCSE assessment comprised two to four units offered on up to four separate occasions. So, with re-sits, it would be possible for a student taking 10 GCSEs to take 70 examinations over the two-year period of study. Vidal Rodeiro and Nádas (2012) found that only two per cent of students re-sat more than three modules but that, dependent upon module, between a third and a half improved their grade by re-sitting. Investment of time and money in re-sitting is significant for pupils, teachers and schools. The examination structure, in particular a modular structure, has the potential to impact students inequitably dependent on resources (such as examination fees) and motivation; different subgroups may be affected differently.

### ***Socio-economic equity gap and GCSE***

There is a wealth of research demonstrating the impact of SES on GCSE outcomes. In a longitudinal study, Sammons et al. (2014) found that parents' highest level of qualification when the child was age 3 or 5 was the strongest predictor of GCSE mathematics outcomes (effect size = 0.74 for higher degree versus no qualification). Prior attainment, measured at the end of primary education, had a higher effect size (2.54), but since educational outcomes at primary level are confounded by SES and other variables, this was analysed separately, as academic progress, by Sammons et al. (2014). Family SES (professional versus unskilled employment) was a strong positive predictor of outcomes too. Children who received free school meals (FSMs) attained lower grades in GCSE mathematics – an effect that was stronger than for GCSE English. Ethnicity (a positive effect of having Indian heritage), home learning environment and institutional effects, such as indicators of school quality, were also important but they are not the focus of the current research. Socio-economic status (occupation of head of household, parents' educational qualifications, FSMs, home ownership, neighbourhood deprivation – IDACI index) was found to have double the effect of ethnicity and six times the effect of gender in another longitudinal study of GCSE examination results (Strand, 2014). Government policy has attempted to address this inequality by provision of Pupil Premium funding to close attainment gaps for disadvantaged pupils. Pupil Premium funding was introduced in 2011 just before the linearisation of GCSEs (Foster & Long, 2017). Given the importance of GCSE mathematics for

life skills and for future study, the SES equity gap is important and any potential change to that gap caused by reforming the examination structure should be evaluated.

Educational commentators often express the view that modular examinations are more suitable for lower performing students, including those with lower SES and those in receipt of FSM (e.g. NASUWT, 2011). We have not been able to identify a research article prior to Baird et al. (2019) that examined the SES equity gap comparing modular and linear examinations. This lack of evidence was remarkable because a major policy driver for the introduction of modular examinations internationally is a widening participation agenda, particularly in higher and vocational education. Credit frameworks in which students are able to accumulate credit towards a qualification have been seen as important emancipatory tools for those returning to study, such as adult learners (e.g. Dearing, 1997). The introduction of modular examinations at GCSE was part of a policy context in which parity of esteem between vocational and academic qualifications was being sought and there was discussion of fully-blown connective forms of modularity at GCSE, incorporated in vocational Diploma examinations (Tomlinson, 2004). Modular examinations were introduced in a policy context which sought to decrease disadvantage through promotion of lifelong learning and personalised learning through a connected, modular, qualifications framework (see Baird et al., 2019).

Baird, et al. (2019) reported gaps in attainment based on SES, gender and centre type for GCSE English, mathematics and science qualifications. Unlike the current study, findings included all students, no matter their age. There was no educationally significant difference in outcomes for the mathematics and science students between the modular and linear examinations, but for GCSE English the modular outcomes were better for more disadvantaged students. A fuller treatment of the mathematics analyses by SES is given here and we also touch upon the findings by gender, though they are not the focus of this article.

### ***Standard setting***

Much of the literature analysing examination equity gaps treats the issue of standards as though it were entirely unproblematic. However, in England in particular, there is a lively literature on the way that standards are set (Baird et al., 2018). Examination structures have indirect effects upon examination outcomes that are mediated by standard setting processes. A key aim of the reform to linear examinations was to increase the demands of GCSEs because there was a view from the 2010 Coalition government that modularity had contributed to grade inflation (Hansard, 17 September 2012 col 653). However, standard setting policy mediates examination outcomes, as students' learning and assessment experiences are not the same as their grading outcomes. Our interest in this article is in comparing modular and linear GCSE outcomes. As the prevalence of modular GCSEs varied across years, we must interpret the outcomes in relation to the standard setting policies being applied, as well as the examination structures. For example, if standards were more bunched due to a very severe standard setting policy, there may be less of an attainment gap. It could then appear that the examination structure was the cause when in fact the standard setting policy also needs to be considered.

Since 2010, Ofqual's regulatory process has required that standards for the GCSE are set using a 'comparable outcomes' approach (Ofqual, 2015).<sup>v</sup> This method takes into account

prior attainment at age 11 (Key Stage 2) to predict GCSE outcomes for a particular examination entry for each syllabus. However, importantly for this research, standard setting is something of a blunt instrument used to ensure qualification-level outcomes are aligned. In modular examinations, staged assessment and re-sitting both add difficulty to the standard setting process because module standards are set before qualification outcomes are available. Module entry patterns, which could be linked to factors such as school resources and student motivation, might therefore influence individual grade outcomes.

## **Aims**

This research monitors the effect of the return to linear examinations after the reforms introduced in 2012 and 2014. It evaluates the following issues over the period 2007 to 2014:

1. The balance of entries for modular and linear examinations;
2. The characteristics of students entered for modular and linear examinations;
3. The difference in outcomes for modular and linear examinations, controlling for the nature of the candidates taking the qualifications and considering, in particular, the effect of SES.

GCSE mathematics is key to academic progression in England, Wales and Northern Ireland. It was one of the few subjects available in both modular and linear form throughout the period analysed. No study to date has considered how modular or linear assessment might be related to SES equity gaps for traditional GCSE mathematics students, i.e. those who are state educated and taking the examination at the age of 16. By considering the largest single group affected by GCSE policy changes, this research adds to the literature concerning equity in education. It can be used, firstly, for monitoring the fairness of the current structures and, secondly, for informing future decisions about examination structures in the event that modularity is reconsidered in England or elsewhere.

## **Data**

The National Pupil Database (NPD) is an archive holding a wide range of information about students who attend schools and colleges in England. It combines student-level examination results with details of student and school characteristics. A subset of these data is used to evaluate the impact of modular and linear assessment on the performance of low SES students in GCSE mathematics. There were 38 GCSE mathematics syllabuses on offer over the period of the study. Of those, 17 were linear, 18 were modular and three switched between linear and modular. Table 1 shows the balance of entries between 2007 and 2014.

### **[TABLE 1 ABOUT HERE]**

The data were filtered to include only students who were state educated and 16 years of age at the end of the academic year. This refinement ensured there were almost no missing data for any of the covariates selected for inclusion in the analysis. However, the few cases with missing data were excluded. As some assessment routes allow certification at more than one time point, only the best certificated attempt was retained. A variable was constructed to indicate for each record whether the GCSE result was obtained from a modular or linear assessment route.



In England, a key benchmark in school performance measures is the percentage of students that exceed the grade C threshold. While this is not the pass/fail threshold, it is also often used to determine students' eligibility for further study. To measure relative success in GCSE mathematics, therefore, a binary contrast was added to indicate that (1) a student gained a grade A\* to C or (0) a student gained a grade D to U.<sup>vi</sup>

Other potential, measurable covariates were coded for inclusion in the analysis. Student prior attainment is measured as the average Key Stage 2 fine points score.<sup>vii</sup> This score takes values between zero and 40 but was normalised for analysis so that differences between scores could be described in terms of standard deviations. So too was the IDACI (Income Deprivation Affecting Children Index). The IDACI measures the proportion of all children below 16 years of age living in income-deprived families at a postcode level. A higher IDACI score denotes a greater level of deprivation in the area. The section of the NPD concerned with examination results only includes postcode at a school level. Therefore, the IDACI is included as a school-level variable. As such, it is a proxy for the collective deprivation of students attending the school. This covariate has limitations as there are cases where schools situated in deprived areas attract privileged students and vice versa. However, not only does it provide an indication of SES, preliminary analysis shows that it is a far stronger school descriptor than any broad centre type classification.

Student gender was coded such that the baseline category was female. Lastly, a binary contrast was created to identify students in receipt of FSMs. Students are eligible for FSMs if they or their guardians are eligible for income support or other such welfare. This variable is often used as a proxy for socio-economic disadvantage, although it is not without its problems (Hobbs & Vignoles, 2010). Firstly, claims must originate from the parent or guardian of the student and so not all that are eligible will claim. Secondly, the variable omits many nuances of socio-economic disadvantage. For modelling, the base category for FSMs was a child *not* in receipt of this benefit.

The year of examination entry was not included as a covariate in the model because, to a certain extent, it was confounded with assessment structure.

## Method

Data included in the NPD are observational in nature and entries for the modular and linear examinations were not random allocations. Raw differences between modular and linear assessment routes could, therefore, be distorted as a result of selection biases in entry patterns. Indeed, Table 2 shows that a higher percentage of students eligible for FSMs were entered for a linear assessment route and that modular students had a slightly higher prior attainment measure. Furthermore, as the modular GCSE mathematics syllabuses were phased out, the cohort of students entered for these syllabuses became less disadvantaged and had higher prior attainment. Vidal Rodeiro and Nádas (2012) also found that modular GCSE mathematics attracted students with higher prior attainment.

### [TABLE 2 ABOUT HERE]

While the use of matching to create quasi-experimental conditions was considered, the limited advantage gained by improving balance in the data was outweighed by an 18%

reduction in the size of the dataset. Additionally, the findings from the matched and unmatched data were equivalent. Therefore, to increase understanding of how modular and linear assessment affects the performance of low SES students in GCSE mathematics, a multilevel logistic model was fitted using the full dataset. This model was chosen to exploit the inherent hierarchy in the data reflecting the fact that a student's GCSE result is a function of both individual and school influences. Indeed, preliminary analysis suggested that over 40% of the variation in students' GCSE mathematics grades could be attributed to the school.

The models were fitted using the lme4 package for R (Bates et al., 2015). The probability of gaining a grade A\* to C was the dependent variable. All educationally relevant covariates were retained regardless of statistical significance. In addition to the main effects, the final model (Model 4, see Table 3 below) included a quadratic term for prior attainment and interactions between the two measures of SES and assessment route (Equation 1). There was no evidence of a non-linear relationship between IDACI and the dependent variable.

$$\log \left( \frac{\text{Prob.A}^*\text{C}_{ij}}{1-\text{Prob.A}^*\text{C}_{ij}} \right) = \beta_{0j} + \beta_{1j} \text{Male}_{ij} + \beta_{2j} \text{Prior.Att}_{ij} + \beta_{3j} \text{Prior.Att}_{ij}^2 + \beta_{4j} \text{IDACI}_{ij} + \beta_{5j} \text{FSM}_{ij} + \beta_{6j} \text{Modular}_{ij} + \\ \beta_{7j} \text{IDACI} * \text{Modular}_{ij} + \beta_{8j} \text{FSM} * \text{Modular}_{ij} + \beta_{9j} \text{IDACI} * \text{FSM}_{ij} + \\ \beta_{10j} \text{IDACI} * \text{FSM} * \text{Modular}_{ij} + u_j \\ \beta_{0j} = \beta_0 + u_{0j} \\ u_{0j} \sim N(0, \sigma_{u0}^2)$$

**Equation 1**

The final model provided a good fit for the data. Addition of the ten covariates resulted in a significant improvement from the null model. The variance explained ( $R^2$ ) by the fixed and random effects was estimated to be 0.65 (Nakagawa & Schielzeth, 2013, equation 30). There was no evidence of over-dispersion in the data.

## Results and discussion

Table 3 shows the results of the models that were fitted. Inevitably, when dealing with a vast number of GCSE mathematics outcomes, the volume of data means that even small differences in outcomes between subgroups of students are statistically significant. Interpretation of this statistical significance and any subsequent inference of educational significance, however, require some unpacking.

With the introduction of complex modelling, the wish to summarise data by means of p-values has given much cause for debate (Bolker, 2017; Bates et al., 2015). Therefore, instead of reporting p-values in the models in Table 3, a † has been placed beside each covariate where the ratio of the parameter estimate to its reported standard error is greater than two. This approximate approach to statistical significance is only really appropriate given the volume of data involved and the central focus on educational impact.

Given the limited evidence provided by any significance testing, the effect size of each covariate is also reported for the final model. Effect sizes for the logistic model are often reported as odds ratios (Rosenthal, 1996). As all of the covariates in the analysis of GCSE mathematics are binary or standardised, direct comparison between quoted odds ratios is valid.

For completeness, Table 3 includes the null model (Model 1). Each subsequent model represents a significant improvement on the null model. Model 2 includes all covariates, and their interactions with SES, but omits prior attainment. Model 3 includes only the main effect covariates and Model 4 adds the interactions with SES.

From Models 3 and 4, it is clear that no matter what a student's background the main factor associated with GCSE mathematics success is prior attainment. A one standard deviation increase in this covariate produces a twelve-fold increase in the odds of exceeding the grade C threshold. Furthermore, the quadratic term indicates that rate of progress between Key Stage 2 assessment and GCSE assessment is greater for the more able students.

However, students' formative experiences affect the measure of prior attainment. While the influence of the other covariates on GCSE mathematics outcomes appears relatively small, this may be because the measure of prior attainment has already been impacted by them. In modelling terms this phenomenon is often referred to as endogeneity bias (Mayston, 2006). Indeed, the 2014 Department for Education press release for national curriculum assessments at Key Stage 2 shows that, even by the end of primary education, students in receipt of FSM have weaker Key Stage 2 outcomes (Department for Education, 2014). After accounting for prior attainment, therefore, the effect size for other covariates should be interpreted as a measure of the impact of secondary education on GCSE outcomes; or a measure of progress (Sammons et al., 2014). The effect sizes in Model 4 reveal only patterns that have emerged since Key Stage 2 assessment; hence the inclusion of Model 2 for comparison. Although the variance explained by Model 2 is only 0.38 compared with 0.65 for Model 4, the fact that it excludes prior attainment means it gives a longer-term measure of the effect of the other covariates.

### **[TABLE 3 ABOUT HERE]**

As discussed, social disadvantage is identified as one of the key limitations to academic success (Sammons et al., 2014; Strand, 2014) and the government response has been to introduce Pupil Premium funding. Over the period of study, the impact of Pupil Premium funding is yet to be felt, and the models described in Table 3 reveal that social disadvantage is still associated with poorer performance in GCSE assessment (Models 2-4). For FSM students and those with high IDACI, the odds of gaining A\* to C is considerably lower than their more socially advantaged counterparts. Furthermore, when prior attainment is excluded from the model (Model 2), this effect is more extreme than when it is included (Model 4). In keeping with previous studies (for example, Feinstein, 2004), the findings show that the current secondary education system exacerbates patterns of inequity already present by the end of primary education.

Perhaps counterintuitively, given the lack of re-sit opportunities, evidence from the models in Table 3 suggests that GCSE mathematics students generally performed better in linear examinations. While this could be interpreted as modular examination being harder, outcomes for modular and linear assessment routes are inextricably intertwined with the standard setting process and with unmeasurable choices that are made in the course of a student's education. GCSE mathematics syllabuses perceived to have high utility for further study might be selected by a school regardless of the teaching and assessment structure. Modular syllabuses may be taught in a linear manner and vice versa. Linear examinations

might be offered to less able students to maximise teaching time; alternatively, the same students might be offered modular examinations to reduce the burden of assessment. Choices might not be open decisions; schools might be limited by resources or by the availability of syllabuses. Furthermore, hidden beneath each modular certification are many unit re-sit opportunities. All of these factors, and more, make it difficult to interpret any differences in outcome between modular and linear assessment routes.

Despite the inherent disadvantage for those with low SES, however, there is very little evidence that assessment route exacerbated the problem. The covariate describing the interaction between FSM and assessment route is negative but, as with all the interactions between FSM, IDACI and assessment route (Model 4), it has a relatively large standard error and correspondingly small effect size. The main effect parameter estimates differ very little between Model 3 and Model 4. Based on Model 4, Figure 1 shows the probability of gaining a grade A\* to C for high SES students compared with low SES students. Low SES has been defined as a student in receipt of FSM and attending school in an area with an IDACI score of one standard deviation above the norm. In contrast, high SES has been defined as a student not in receipt of FSM and attending school in an area with an IDACI score of one standard deviation below the norm. The outcomes for a modular assessment route are lower regardless of SES. The outcomes for a low SES student are lower regardless of assessment route. But the gap between the more advantaged and more disadvantaged students is no greater for modular assessment than for linear assessment. In other words, the linearisation of GCSE mathematics assessments is unlikely to have exacerbated existing social disadvantage.

**[FIGURE 1 ABOUT HERE]**

## **Conclusions**

In keeping with policy changes, by 2014 nearly all GCSE mathematics entries were linear (99%). Entries for modular examinations were balanced between genders. However, as the modular examinations were phased out, prior attainment became slightly higher for modular examinations and eligibility for FSMs became lower indicating that, latterly, a more advantaged group were entered for modular GCSE mathematics examinations.

The results of the current analyses across 2007-2014 indicated that after controlling for prior attainment, gender and indicators of disadvantage (FSM and IDACI), modular examination candidates were likely to have worse outcomes. Thus, it is possible that modular GCSE mathematics examinations were graded severely, statistically speaking. This is a surprising finding, as it runs counter to the policy rationale for the linearisation of GCSEs; that modular examinations were too easy. However, the policy intentions for the introduction of linear examinations also incorporated the re-sit culture, domination of examinations over children's educational experiences and multiple entries for the same examinations. Further, the grading of modular GCSE mathematics may have appeared more severe in these statistical analyses, but it is possible that this was deserved, due to the quality of candidates' performances in the examinations. That is, the modular candidates may have been less good at mathematics and the standard setting system may have identified these differences. Preparation for modular GCSEs may have been more disruptive to mathematics progress between the prior attainments at age 11 and the GCSEs, taken at

age 16. Thus, ascertaining whether the modular examinations were ‘dumbed down’ in relation to student performances would require a broader range of research methods than presented in this article. However, these findings show that the standard setting process ensured they were not dumbed down in relation to GCSE mathematics outcomes.

The results from this research show that, in GCSE mathematics, there is a clear gap in attainment between disadvantaged students and their peers. However, this gap did not seem to be exacerbated by examination structure and so the linearisation of GCSE mathematics is unlikely to further increase inequity between students of high and low SES. Access to the highest GCSE grades is dependent upon teachers entering students for the higher tier question papers. Vitello & Crawford (2018) showed that for all but one GCSE mathematics syllabus assessment structure (modular or linear) had no effect on the choice of entry tier. However, they did find weak evidence that level of deprivation interacted significantly with the assessment structure to affect the probability of being entered into the foundation tier. Teachers’ examination entry patterns may be the result of bias or informed by students’ performances.

The picture that emerges is complex and appears to differ between subjects (see Baird et al., 2019). It is unlikely that linearisation has had a disproportionate impact upon disadvantaged students but neither has it reduced the attainment gap.

### ***Limitations***

Outcomes for modular and linear assessment routes are intertwined with the standard setting process. No standard setting process is infallible and weaknesses of the ‘comparable outcomes’ approach to standard setting – in which prior attainment at age 11 is used to predict GCSE outcomes for a particular examination entry for each syllabus (Ofqual, 2015) – complicate interpretation of the findings; specifically, the issue of non-random entry to routes as reported by Taylor (2016).

Taylor (2016) conducted semi-structured interviews with secondary school teachers in England to explore how the pressure of school performance measures impact behaviour in schools. One of the strategies reported involved switching students between modular and linear GCSEs. Some students were entered for modules throughout their course of study. If, based on their module results, they were likely to achieve grade C or above they were entered for their final module and certificated. If, on the other hand, they were unlikely to achieve grade C or above they were switched to the linear GCSE to allow a fresh attempt at all the assessments comprising the linear GCSE.

The ‘comparable outcomes’ approach is undermined by this systematic practice. Students who underperformed relative to their prior attainment were moved from the modular to the linear GCSE meaning that using prior attainment to inform boundary setting would lead to boundaries that were too low for the linear GCSE and boundaries that were too high for the modular GCSE. While such issues were identified at the time of standard setting it is difficult to estimate the size of such effects and so to fully control for them. Moreover, use of early and multiple examination entry varies between schools and will be partly determined by resource availability. Schools with relatively higher proportions of disadvantaged students might be more motivated to use such tactics to maximise results

but they might have fewer resources to do so. It is likely that the use of such practices, route entry choices, standard setting and student disadvantage interact such that we must be cautious in drawing firm conclusions from this work. In addition, this research only investigated data from GCSE mathematics and it is possible that different effects would be found for other comparisons of examination structure for different subjects, examinations and standard setting procedures.

## **Acknowledgements**

The project was conducted collaboratively between Oxford University Centre for Educational Assessment and Ofqual and is funded by Ofqual.

## **Data Availability Statement**

The National Pupil Database data that support the findings of this study are available from the Department for Education. Restrictions apply to the availability of these data, which were used under license for this study. The modular flag variable is available from Ofqual.

## **Ethical Guidelines**

Ethical approval for this study was gained from the University of Oxford, using BERA guidelines through the Department of Education.

## **Conflicts of Interest**

The authors certify that they have no interests or affiliations that may influence, or be perceived to influence, the objectivity of the findings discussed in this manuscript.

## **References**

- Baird, J., Caro, D., Elliott, V., El Masri, Y., Ingram, J., Isaacs, T., Pinot de Moira, A. Randhawa, A., Stobart, G., Meadows, M., Morin, C., & Taylor, R. (2019). *Examination Reform: Impact of Linear and Modular Examinations at GCSE*. Ofqual/19/6506/1. Retrieved from <https://www.gov.uk/government/publications/examination-reform-impact-of-linear-and-modular-examinations-at-gcse>.
- Baird, J., Isaacs, T., Opposs, D., & Gray, L. (Eds.) (2018). *Examination standards: how measures and meanings differ around the world*. London: UCL IOE Press.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bolker, B. (2017). *GLMM FAQ: Inference and confidence intervals*. Retrieved from <http://bbolker.github.io/mixedmodels-misc/glmmFAQ.html#inference-and-confidence-intervals>.
- D’Arcy, J. (Ed.) (1997). *Comparability studies between modular and non-modular syllabuses in GCE Advanced Biology, English Literature and Mathematics in the summer 1996 examination*. Standing Committee on Research on behalf of the Joint Forum for the GCSE and GCE.

- Dearing, R. (1996). *The Dearing Review: Review of Qualifications for 16-19 year olds*. London: School Curriculum and Assessment Authority.
- Dearing, R. (1997). Higher Education in the learning society. London: Her Majesty's Stationery Office. Retrieved from <http://www.educationengland.org.uk/documents/dearing1997/dearing1997.html>.
- Department for Education (2010). *Participation in education, training and employment by 16 - 18 year olds in England. Statistical First Release SFR18/2010, 22 June*. Retrieved from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/218900/main\\_20text\\_20sfr182010.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/218900/main_20text_20sfr182010.pdf).
- Department for Education (2013). *NEET Statistics - Quarterly Brief - Quarter 4 2012. Statistical First Release SFR09/2013, 28 February*. Retrieved from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/219324/sfr09-2013.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/219324/sfr09-2013.pdf).
- Department for Education (2014). *National curriculum assessments at Key Stage 2 in England, 2014 (Revised). Statistical First Release SFR 50/2014, 11 December*. Retrieved from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/428838/SFR50\\_2014\\_Text.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/428838/SFR50_2014_Text.pdf).
- Department for Education (2017). *Education and Training Statistics for the United Kingdom 2017. Statistical First Release SFR64/2017, 9 November*. Retrieved from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/657821/SFR64\\_2017\\_Text.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/657821/SFR64_2017_Text.pdf).
- Ertl, H., & Hayward, G. (2010). Modularization in Vocational Education and Training. In: P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education* (3rd Edition) (pp. 383–390). Oxford: Elsevier. Retrieved from <https://doi.org/10.1016/B978-0-08-044894-7.00786-7>.
- Feinstein, L. (2004). Mobility in Pupils' Cognitive Attainment During School Life. *Oxford Review of Economic Policy*, 20(2), 213–229.
- Foster, D., & Long, R. (2017). *The Pupil Premium: Briefing Paper Number 6700. 12 December*. London: House of Commons Library. Retrieved from <http://dera.ioe.ac.uk/30693/1/SN06700%20.pdf>.
- Gill, T., & Williamson, J. (2016). *Uptake of GCSE subjects 2015*. Statistics Report Series No. 107. Cambridge: Cambridge Assessment. Retrieved from <http://www.cambridgeassessment.org.uk/Images/307016-uptake-of-gcse-subjects-2015.pdf>.
- Hansard HC Deb vol 550 col 653 (17 September 2012).
- Hansard HC Deb vol 550 col 655 (17 September 2012).

Hansard HC Deb vol 558 col 441 (7 February 2013).

Hobbs, G., & Vignoles, A. (2010). Is children's free school meal 'eligibility' a good proxy for family income? *British Educational Research Journal*, 36(4), 673–690.

Leckie, G., & Goldstein, H. (2016). The evolution of school league tables in England 1992-2016 : 'contextual value-added', 'expected progress' and 'progress 8'. Bristol Working Papers in Education Series, #02/2016. Retrieved from [https://www.bristol.ac.uk/media-library/sites/education/documents/bristol-working-papers-in-education/The evolution of school league tables in England 1992-2016.pdf](https://www.bristol.ac.uk/media-library/sites/education/documents/bristol-working-papers-in-education/The%20evolution%20of%20school%20league%20tables%20in%20England%201992-2016.pdf).

Mayston, D. (2006). *Educational value added and programme evaluation*. Retrieved from <http://webarchive.nationalarchives.gov.uk/20130321025016/https://www.education.gov.uk/publications/eOrderingDownload/RW87.pdf>.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining  $R^2$  from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142.

NASUWT (2011). *Symposium on Qualifications Reform in England*. Birmingham: NASUWT. .

Ofqual (2015). *Setting GCSE, AS and A Level Grade Standards in summer 2014 and 2015*. Ofqual/15/5759. Coventry: Office of Qualifications and Examinations Regulation. Retrieved from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/451321/2015-08-05-summer-series-gcse-as-and-a-level-grade-standards.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/451321/2015-08-05-summer-series-gcse-as-and-a-level-grade-standards.pdf).

Qualifications and Curriculum Authority (2004). *GCSE, GCSE in vocational subjects, GCE, VCE, GNVQ and AEA Code of Practice 2004/5*. QCA/04/1282. London: Qualifications and Curriculum Authority. Retrieved from [http://dera.ioe.ac.uk/6264/1/6295\\_gcse\\_vocgcse\\_gce\\_vce\\_gnvq\\_aea\\_code-of-practice\\_04-05.pdf](http://dera.ioe.ac.uk/6264/1/6295_gcse_vocgcse_gce_vce_gnvq_aea_code-of-practice_04-05.pdf).

Qualifications and Curriculum Authority (2006). *Review of standards in A level computing and ICT: 1998–2004*. QCA/06/2346. London: Qualifications and Curriculum Authority. Retrieved from <http://dera.ioe.ac.uk/8917/1/qca-06-2346-computing-a-level.pdf>.

Qualifications and Curriculum Authority (2007). *A level resitting: summary of research findings*. QCA/07/3387. London: Qualifications and Curriculum Authority. Retrieved from <http://dera.ioe.ac.uk/8669/1/qca-07-3387-Resit-report.pdf>.

Qualifications and Curriculum Authority (2008). *Regulatory arrangements for the Qualifications and Credit Framework*. London: Qualifications and Curriculum Authority. Retrieved from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/371294/2008-08-15-regulatory-arrangements-qcf-august08.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/371294/2008-08-15-regulatory-arrangements-qcf-august08.pdf).

Rosenthal, J. A. (1996). Qualitative Descriptors of Strength of Association and Effect Size.



*Journal of Social Service Research*, 21(4), 37–59.

- Sammons, P., Sylva, K., Melhuish, E., Siraj, I., Taggart, B., Toth, K., & Smees, R. (2014). *Influences on students' GCSE attainment and progress at age 16. Effective Pre-School, Primary & Secondary Education Project (EPPSE) Research Report*. London: Institute of Education, University of London. Retrieved from [http://dera.ioe.ac.uk/20875/1/RR352\\_-\\_Influences\\_on\\_Students\\_GCSE\\_Attainment\\_and\\_Progress\\_at\\_Age\\_16.pdf](http://dera.ioe.ac.uk/20875/1/RR352_-_Influences_on_Students_GCSE_Attainment_and_Progress_at_Age_16.pdf).
- Strand, S. (2014). Ethnicity, gender, social class and achievement gaps at age 16: intersectionality and 'getting it' for the white working class. *Research Papers in Education*, 29(2), 131–171.
- 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.
- Tomlinson, M. (2004). 14-19 Curriculum and Qualifications Reform. Final Report of the Working Group on 14-19 Reform. October. Department for Education and Skills Publications. DfE-0976-2004. Retrieved from <http://webarchive.nationalarchives.gov.uk/20050301194752/http://www.dfes.gov.uk/14-19/documents/Final%20Report.pdf>.
- Vidal Rodeiro, C. L., & Nádas, R. (2012). Effects of modularity, certification session and re-sits on examination performance. *Assessment in Education: Principles, Policy & Practice*, 19(4), 411–430.
- Vitello, S., & Crawford, C. (2018). Which tier? Effects of linear assessment and student characteristics on GCSE entry decisions. *British Educational Research Journal*, 44(1), 94–118.

---

<sup>i</sup> <https://www.gov.uk/government/publications/get-the-facts-gcse-and-a-level-reform/get-the-facts-gcse-reform>

<sup>ii</sup> <https://www.gov.uk/government/news/setting-standards-for-new-gcses-in-2017>

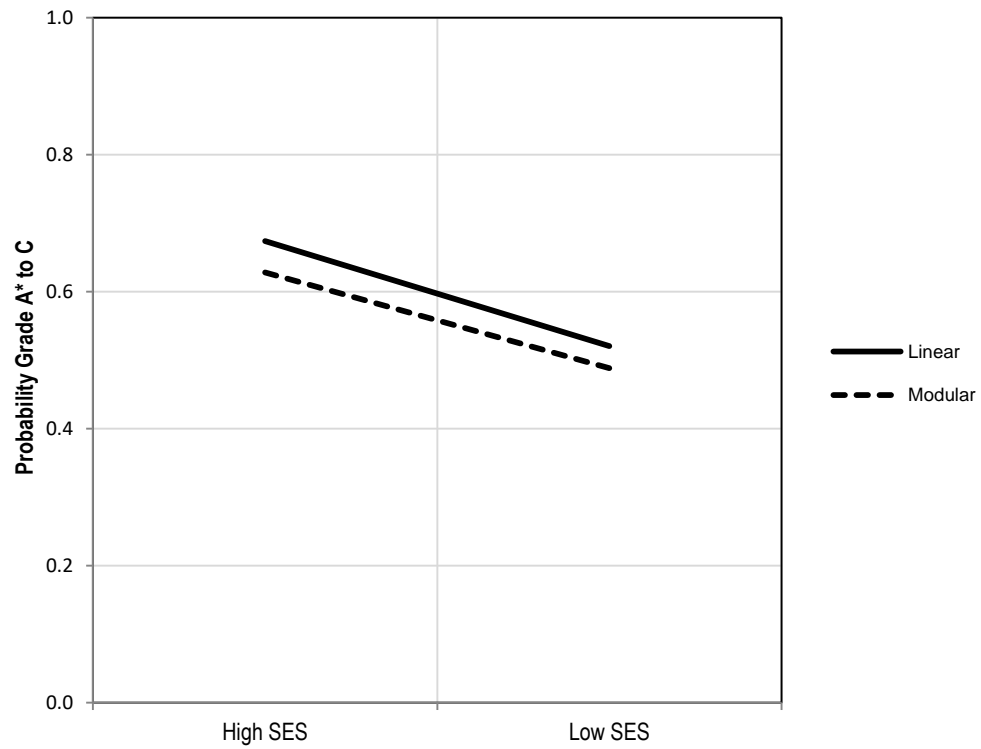
<sup>iii</sup> <https://www.gov.uk/government/publications/english-baccalaureate-ebacc/english-baccalaureate-ebacc>

<sup>iv</sup> <https://www.gov.uk/government/publications/progress-8-school-performance-measure>

<sup>v</sup> Between 2007 and 2010, an early version of the comparable outcomes approach was in place across examination boards.

<sup>vi</sup> Since the move towards linear assessment, all GCSEs in England have undergone further reforms. Amongst other things, these reforms include a change to the grading scale where, from summer 2017, 9 became the highest grade and 1 the lowest.

<sup>vii</sup> The Key Stage 2 tests discussed here were taken by students at the age of 11: the end of primary school. They were assessed in reading, writing, spelling, punctuation, grammar and mathematics. Test marks were combined to create a composite score. These tests were revised in 2014 and are no longer graded in the same way.



**Table 1. The balance of modular and linear GCSE Mathematics entries in the academic years 2006/07 to 2013/14**

	2007	2008	2009	2010	2011	2012	2013	2014	Total
Linear (%)	61.8	55.0	52.1	52.4	57.2	67.3	77.8	98.9	65.0
Modular (%)	38.2	45.0	47.9	47.6	42.8	32.7	22.2	1.1	35.0
Entries (000s)	538	538	520	521	511	495	509	495	4,126

**Table 2. Entry characteristics for modular and linear assessment routes**

		2007	2008	2009	2010	2011	2012	2013	2014
Male (%)	Linear	50.19	50.45	50.26	50.10	50.45	50.89	50.61	50.62
	Modular	50.09	50.26	50.48	50.76	50.55	50.10	50.33	51.12
FSM (%)	Linear	11.70	11.63	12.26	13.45	14.89	16.06	16.29	14.26
	Modular	12.14	11.71	11.96	12.01	11.46	9.45	8.48	7.09
Prior	Linear	27.47	27.41	27.48	27.46	27.42	27.24	27.34	27.72
Attainment (Mean)	Modular	27.01	27.06	27.27	27.61	28.11	28.69	29.24	30.02
IDACI	Linear	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.17
(Mean)	Modular	0.17	0.17	0.17	0.17	0.16	0.15	0.14	0.14

**Table 3. Parameter estimates for the multilevel logistic regression analysis of the probability of gaining a grade A\*-C**

Fixed Effects	Model 1 (Null)			Model 2			Model 3			Model 4 (Final)		
	Estimate	Std. Err	Odds	Estimate	Std. Err	Odds	Estimate	Std. Err	Odds	Estimate	Std. Err	Odds
Intercept	0.175	0.023	1.191	0.308	0.022 <sup>†</sup>	1.361	0.586	0.015 <sup>†</sup>	1.796	0.586	0.015 <sup>†</sup>	1.798
Gender: Male				0.017	0.002 <sup>†</sup>	1.017	0.161	0.003 <sup>†</sup>	1.175	0.161	0.003 <sup>†</sup>	1.175
Prior Attainment							2.477	0.003 <sup>†</sup>	11.907	2.477	0.003 <sup>†</sup>	11.905
Prior Attainment ^ 2							0.345	0.002 <sup>†</sup>	1.412	0.345	0.002 <sup>†</sup>	1.412
IDACI				-0.121	0.010 <sup>†</sup>	0.886	-0.106	0.010 <sup>†</sup>	0.900	-0.139	0.010 <sup>†</sup>	0.871
FSM: Yes				-0.837	0.004 <sup>†</sup>	0.433	-0.429	0.004 <sup>†</sup>	0.651	-0.453	0.005 <sup>†</sup>	0.636
Modular				0.064	0.003 <sup>†</sup>	1.066	-0.162	0.004 <sup>†</sup>	0.851	-0.161	0.004 <sup>†</sup>	0.851
IDACI * Modular				0.016	0.003 <sup>†</sup>	1.016				0.042	0.004 <sup>†</sup>	1.043
FSM: Yes * Modular				-0.071	0.007 <sup>†</sup>	0.931				-0.017	0.009	0.983
IDACI * FSM: Yes				0.130	0.004 <sup>†</sup>	1.139				0.086	0.005 <sup>†</sup>	1.090
IDACI * FSM: Yes * Modular				-0.007	0.006	0.993				0.008	0.008	1.008
<b>Random Effects</b>	<b>Variance</b>	<b>Std. Dev</b>		<b>Variance</b>	<b>Std. Dev</b>		<b>Variance</b>	<b>Std. Dev</b>		<b>Variance</b>	<b>Std. Dev</b>	
School residual	2.306	1.518		2.069	1.438		0.885	0.941		0.880	0.838	
<b>R<sup>2</sup></b>	<b>0.412</b>			<b>0.397</b>			<b>0.649</b>			<b>0.649</b>		

<sup>†</sup> Denotes estimate is more than twice the standard error

