

MAKING THE GRADE

The impact of GCSE reforms on the attainment gap
between disadvantaged pupils and their peers

Simon Burgess and Dave Thomson
– December 2019





About the Sutton Trust

The Sutton Trust is a foundation which improves social mobility in the UK through evidence-based programmes, research and policy advocacy.

Copyright © 2019 Sutton Trust. All rights reserved. Please cite the Sutton Trust when using material from this research.

About the authors

Simon Burgess is Professor of Economics at the University of Bristol. His current research interests are in the economics of education, including the importance of teachers, pupils and schools, market structure in education, incentives, and choice. He has also worked on the performance of ethnic minorities in schools and ethnic segregation.

Dave Thomson is chief statistician at FFT. He has over fifteen years' experience working with educational attainment data to raise attainment in local government, higher education and the commercial sector. His current research interests include linking education and workplace datasets to improve estimates of adult attainment and studying the impact of education on employment and benefits outcomes.

Contents

Executive Summary	3
Introduction	4
<i>GCSEs are important</i>	4
<i>Context: the reforms and the reactions</i>	5
Methodology	7
<i>Empirical methods</i>	7
<i>The National Pupil Database</i>	8
<i>Outcomes</i>	9
GCSE reforms and the attainment gap	13
<i>Main pupil-level results</i>	14
<i>Attainment at key grades</i>	16
Changes by subject	20
<i>Attainment in English and mathematics</i>	20
<i>Variation between subjects</i>	22
Summary and conclusions	23

Executive Summary

- Reforms to GCSEs were introduced in 2015, with the first cohorts taking the new exams in 2017 and 2018 across a range of subjects. The major changes comprised of more challenging material, a move from modular assessment to a focus on final exams, and a change in the grading system from letters (A*, A...G etc), to numbers (9, 8...1). The aims of the reforms were to improve standards overall by making courses harder and increase differentiation by making the top grades rarer.
- Pupils are overall entering slightly fewer GCSEs since the reforms, but the gap in entries between disadvantaged students and others has decreased.
- There has been a slight increase in those achieving grades 9-4 (equivalent to the old A*-C), for both disadvantaged and non-disadvantaged pupils. Though there is little perceptible difference at the top end; those achieving grades 9-7 (A*-A). For high achieving pupils at the end of primary school, the patterns are similar. Any changes are mainly for middle achievers.
- During the period of the reforms, there has been a small impact on increasing the disadvantage attainment gap. Accounting for a variety of school factors and pupil characteristics, test scores for disadvantaged pupils fell slightly compared to non-disadvantaged pupils, by just over a quarter of a grade across 9 subjects (1.5% of a standard deviation per subject).
- While differences in the attainment gap at the A/7 and C/4 boundaries are minimal, at least partially owing to the deliberate maintenance of grade boundaries for consistency over time, there are potential impacts at other grade levels. Under the previous system, 2% of disadvantaged pupils achieved the top grade of A*, whereas just 1% now achieve a 9. The drop is less for non-disadvantaged pupils, falling from 8% achieving A* to 5% achieving a 9. Similarly, grade 5 is now seen as a 'strong pass'. Non-disadvantaged pupils were 1.42 times more likely than disadvantaged pupils to achieve a C or above, but are now 1.63 times more likely to achieve a 5 or above.
- Achieving a 'standard pass' in English and maths is crucial to educational progression after age 16, with those who do not reach this level compelled to take resits. There has been a drop of 1 percentage point in the proportion of disadvantaged pupils failing to meet that standard in both English (41.1% in 2016, 40.1% in 2018) and maths (49.8% in 2016, 48.8% in 2018).
- Across most subjects, attainment gaps increased only slightly or did not change, with only the triple sciences showing a more appreciable increase in the attainment gap.
- In sum, substantial increases in the disadvantage attainment gap have not materialised, and this is likely to be in part due to the 'comparable outcomes' approach and the deliberate maintenance of grade boundaries.
- In terms of wider consequences, slight falls in those who do not pass English and maths GCSEs could, if the trend continues, have a positive impact on the post-16 options of disadvantaged pupils at the lower end of the scale, who may otherwise be forced to re-sit. However, greater differentiation at the top end of the ability scale may have negative social mobility impacts, for instance where employers or universities focus on those achieving top marks.

Introduction

GCSEs are important

The English education system has a tradition of public examination at age 16. Since 1987, this has principally been in the form of General Certificates of Secondary Education (GCSE). Although some might now consider them an anachronism in an age of participation until age 18, they remain key because they act as ‘gatekeeper’ to post-16 education routes and success in the job market. Good performance in GCSEs is an important first step in securing strong life chances. Substantial reform of many aspects of GCSE qualifications – content, structure and assessment all together – therefore has the potential to make significant differences to the future prosperity of generations of pupils. The reforms, discussed below, were designed to increase the rigour of the assessment; this had a number of aspects but was achieved chiefly by making the content harder, and by changing the grading system to differentiate more between pupils with high marks.

As in many countries, there is a large and ever-present gap in performance between pupils coming from affluent backgrounds and those from poorer families. Official figures for 2018 show that 44.5% of disadvantaged pupils achieved grades 9-4 in both GCSE English and maths compared to 71.5% of other pupils.² However, this gap appeared to widen in 2018 having closed in previous years.³ Given the gatekeeper role of GCSEs, this attainment gap has consequences for pupils’ subsequent careers by reducing chances for good jobs and higher and further education.⁴

In this report we examine how the major reforms to GCSEs have affected the GCSE attainment gap,⁵ particularly among pupils with high levels of prior attainment. Exams are graded in England adhering to a standard called “comparable outcomes” which aims to ensure continuity in grades when qualifications change.⁷ This means that year to year fluctuations in grades awarded are minimised, so we do not expect to see dramatic changes. Nevertheless, comparing grades between sub-groups is meaningful and that is what we report below. We use examination data covering the population of pupils taking GCSEs in state-maintained mainstream schools over three years straddling the reform. We examine performance at pupil level and at pupil-by-subject level.

The overall result is clear: we find a statistically well determined effect, small but going in the direction of further disadvantaging the disadvantaged. So far at least, and although it could be argued that positive effects may take longer to come through, the GCSE reforms have widened the attainment gap as young people move into the labour market or on to further study.

² <https://www.gov.uk/government/statistics/key-stage-4-and-multi-academy-trust-performance-2018-revised>

³ <https://epi.org.uk/publications-and-research/annual-report-2019/>

⁴ Crawford et al (2017) Family Background and University Success. Oxford: OUP.

⁵ Two papers have looked at this specific case. First, a report commissioned by Ofqual looked specifically at the effects of GCSE reform, and found no changes

(https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/798046/Modular_linear_report_final.pdf). Second, Ofqual have themselves shown that the school-level variability after the GCSE reforms was essentially the same as before (<https://analytics.ofqual.gov.uk/apps/2018/GCSE/CentreVariability/>).

⁷ <https://ofqual.blog.gov.uk/2018/01/19/gcse-and-a-level-awarding-in-2018/>

Context: the reforms and the reactions

The reforms

The reforms were introduced by Michael Gove as Secretary of State for Education, first affecting actual exam outcomes in 2017 (courses starting in 2015); more detail on timing is given below. The main components⁸ are:

- Content – the material taught was to be more challenging, with new content to stretch more able pupils, whilst “remaining accessible” for pupils of all abilities;
- Structure – all the exams were to be at the end of the course (“linear” structure) rather than a more modular approach with exams spread out through the course;
- Assessment – almost entirely by exams, with non-examination assessment only where it was absolutely necessary (for example, assessment of speaking in modern foreign languages);
- Grading – a switch from a system based on letter grades {A*, A, ...G, U} to one based on numbers {9, 8, ...1}. While this may seem innocuous, simple rebranding to distinguish from the old system, it has a number of implications, which we explore further below.
- Tiering – that differentiated tiers, foundation and higher, were permitted only in maths, statistics, science and modern foreign languages.

The aims of the reforms were set out in a letter from Gove to OFQUAL in 2013.¹⁰ These include:

- raising educational standards to match those elsewhere: “there is an urgent need for reform, to ensure that young people have access to qualifications that set expectations that match and exceed those in the highest performing jurisdictions.”, which Gove felt was not the case at the time;
- making the top grades rarer and so more meaningful and valuable: “I am clear that the value of the qualifications for individuals must take precedence ahead of ensuring the absolute reliability of the assessment.”;
- making the courses harder: “At the level of what is widely considered to be a pass ... there must be an increase in demand [i.e. more challenging content], to reflect that of high-performing jurisdictions.”;
- providing a better grounding for further study: “At the top end the new qualification should prepare pupils properly to progress to A levels or other study. This should be achieved through a balance of more challenging subject content and more rigorous assessment structures.”

Overall, the key factors are harder content and more rigorous assessment.

Mechanisms: what differences might the reforms make?

There may be in principle impacts on GCSE outcomes from the new grading system, from the more challenging content or from the different assessment structure.

⁸

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/800993/OfqualDfESlidepack_2019.pdf

¹⁰ https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/529404/2013-02-07-letter-from-michael-gove-reform-of-ks4-qualifications.pdf

The new grading system has the potential for affecting the GCSE attainment gap in a simple mechanical way. For example, consider splitting the old A*-A grades into new grades 7, 8 and 9: if the disadvantaged pupils were disproportionately near the lower end of the A bracket, then the attainment gap at grade 9 could be dramatically higher than that at grade A*. This is not about behaviour change but simply about the distribution of scores within grade boundaries. We see some evidence of this below.

There may also be behavioural responses to the nature of assessment and to the change in content, both in the name of "greater rigour". First, if children from disadvantaged families do tend to score lower overall, making the material harder may interact with that to increase the score gap. Second, if the new material or the new assessment lead some students to struggle, it seems likely that more affluent families will be better placed to spend resources to help their child through it.

Methodology

Empirical methods

We examine different metrics of the GCSE score gap between pupils from disadvantaged and more affluent backgrounds and compare that score gap before and after the reform. The set-up therefore resembles a “difference in difference” approach: estimating whether the reform (before/after) has a different effect on test scores of disadvantaged and other pupils.

This was a time of much reforming activity in secondary school assessment and accountability. Other noteworthy reforms to qualifications and accountability included: the Wolf Review¹¹ (2011), which changed the set of qualifications that schools offered;¹² and the new school accountability system, which focussed on a value-added measure ‘Progress 8’ (Note that this reform was explicitly linked to the reform of GCSEs in Gove’s letter referenced above¹⁴). The key dates of these policies were:

- The English Baccalaureate (EBacc) from 2011, a performance measure to incentivise take-up of GCSEs in English, mathematics, science, humanities (geography and history) and modern and ancient languages, which were believed to “keep young people’s options open for further study and future careers”;¹⁵
- Wolf Review: from 2014, the Key Stage 4 performance tables would be restricted to qualifications that were “high quality, rigorous and enable[d] progression”;
- Progress 8: Progress 8 was first nationally published in January 2017, summarising schools’ performance in the 2015/16 academic year.

We therefore deliberately choose 2016 as the start for our analysis as the Wolf reform and the Progress 8 accountability reform had already taken effect.

One factor that helps us to identify the specific effect of the GCSE reforms is that they were not all reformed in the same year; as noted above, English, maths were changed first (exams in 2017), followed a year later by sciences and most others, and finally some small-scale subjects later still.

It is possible that our results are contaminated by longer-term effects of the earlier policies mingled with the effects of the GCSE reforms, but this is the best that can be done until more years of data become available.

Timing of GCSE Reform

The timing of subject reforms is set out in Table 1. The first set of results in reformed GCSEs in mathematics, English language and English literature were in 2017. The first set of reformed results came through in 2018 for a number of other subjects, including the EBacc subjects.

¹¹ <https://www.gov.uk/government/publications/review-of-vocational-education-the-wolf-report>

¹² <https://onlinelibrary.wiley.com/doi/abs/10.1002/berj.3515>

¹⁴ https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/529404/2013-02-07-letter-from-michael-gove-reform-of-ks4-qualifications.pdf

¹⁵ <https://www.gov.uk/government/publications/english-baccalaureate-ebacc/english-baccalaureate-ebacc>

Table 1: Year first reformed GCSEs were taken

2017	English Language	English Literature	Mathematics
2018	Art & Design	Art & Design (Photography)	Drama
	Dance	Music	Physics
	Biology	Chemistry	Food Preparation and Nutrition*
	Science: Double Award	Computing	Religious Studies
	Geography	History	Spanish
	French	German	
	Latin	Greek	
	Citizenship	Physical Education	

* This was a new subject that overlapped in content to some extent with legacy GCSEs in food technology and home economics.

The National Pupil Database

We use the National Pupil Database (NPD), an administrative dataset covering every pupil in England in state-maintained schools. We use three years of GCSE (Key Stage 4) data (2016 to 2018) linked to the pupil-level school census and prior attainment at Key Stage 2 (age 11). Our analysis is based on pupils at the end of Key Stage 4 attending state-funded schools and who contribute to national results.

Definitions

Our focus is the comparison between disadvantaged pupils and non-disadvantaged pupils; these are defined as “pupils eligible for free school meals (FSM) at any point in the six years up to and including the year in which they reached the end of Key Stage 4”. We see a small drop in the percentage of disadvantaged pupils across the three cohorts (due to changes in the economy and change to eligibility criteria) from 27.1% in 2016 to 26.3% in 2018 (Table 2).

Within the group defined as disadvantaged, we are particularly interested in pupils with high levels of prior attainment. While there is an official definition of “high prior attainment” as the highest of three bands of Key Stage 2 performance, these official data do not account for changes to subjects tested in the KS2 accountability framework in 2012.¹⁶ We therefore redefine the prior attainment bands for the 2017 and 2018 cohorts such that the overall percentage of pupils in the high (and low) groups is broadly similar to 2016, with 29% defined as high attainers. However, among the disadvantaged group, only 16-17% are high attainers (Table 3). Nevertheless, disadvantaged pupils are slightly more likely to be higher attainers in 2018 (17.4%) than in 2016 (15.8%) whereas the rate among the non-disadvantaged stays much the same. So the 2018 cohort of disadvantaged pupils are relatively stronger than the previous cohort, which may introduce a composition effect to our results: this might suggest that the disadvantaged pupils might wrongly appear to do better after the reform.

Pupils classified as disadvantaged vary in the length of time they have been eligible for free school meals during their school career. Some may have only had a brief period of eligibility while others may have always been eligible. This variation matters for subsequent attainment.¹⁷ The three cohorts of

¹⁶ Key Stage 2 tests in writing were abandoned after 2011.

¹⁷ <https://ffteducationdatalab.org.uk/2017/07/long-term-disadvantage-part-one-challenges-and-successes/>

disadvantaged pupils analysed here are broadly similar in terms of length of time eligible for FSM. As our interest is in the group of disadvantaged pupils with higher prior attainment we do not pursue this further here.

Table 2: Number of Disadvantaged Pupils by Cohort (thousands)

	End of Key Stage 4 in		
	2016	2017	2018
Disadvantaged pupils	146.5	139.1	137.3
Other pupils	393.3	387.5	385.1
All pupils	539.8	526.6	522.4
% disadvantaged	27.1%	26.4%	26.3%

Table 3: Number of disadvantaged pupils with higher levels of prior attainment (thousands)

	Number with high PA			High PA as % of cohort		
	2016	2017	2018	2016	2017	2018
Yes	23.1	22.7	23.9	15.8%	16.3%	17.4%
No	134.4	130.5	132.5	34.2%	33.7%	34.4%
All pupils	157.6	153.2	156.4	29.2%	29.1%	29.9%
% disadvantaged	14.7%	14.8%	15.3%			

Outcomes

We focus on attainment in GCSE courses, across the subjects that pupils enter.

Over the reform period, entries in some subjects increased (such as English literature) but fell in others (such as religious studies) but differences in entry rates between disadvantaged pupils and other pupils remained broadly stable. This is most easily seen in Figures 1 and 2. Further data is presented in Appendix Tables A7 and A8.

Figure 1: Percentage of pupils entered for each subject by disadvantage, all pupils in state-funded schools in England

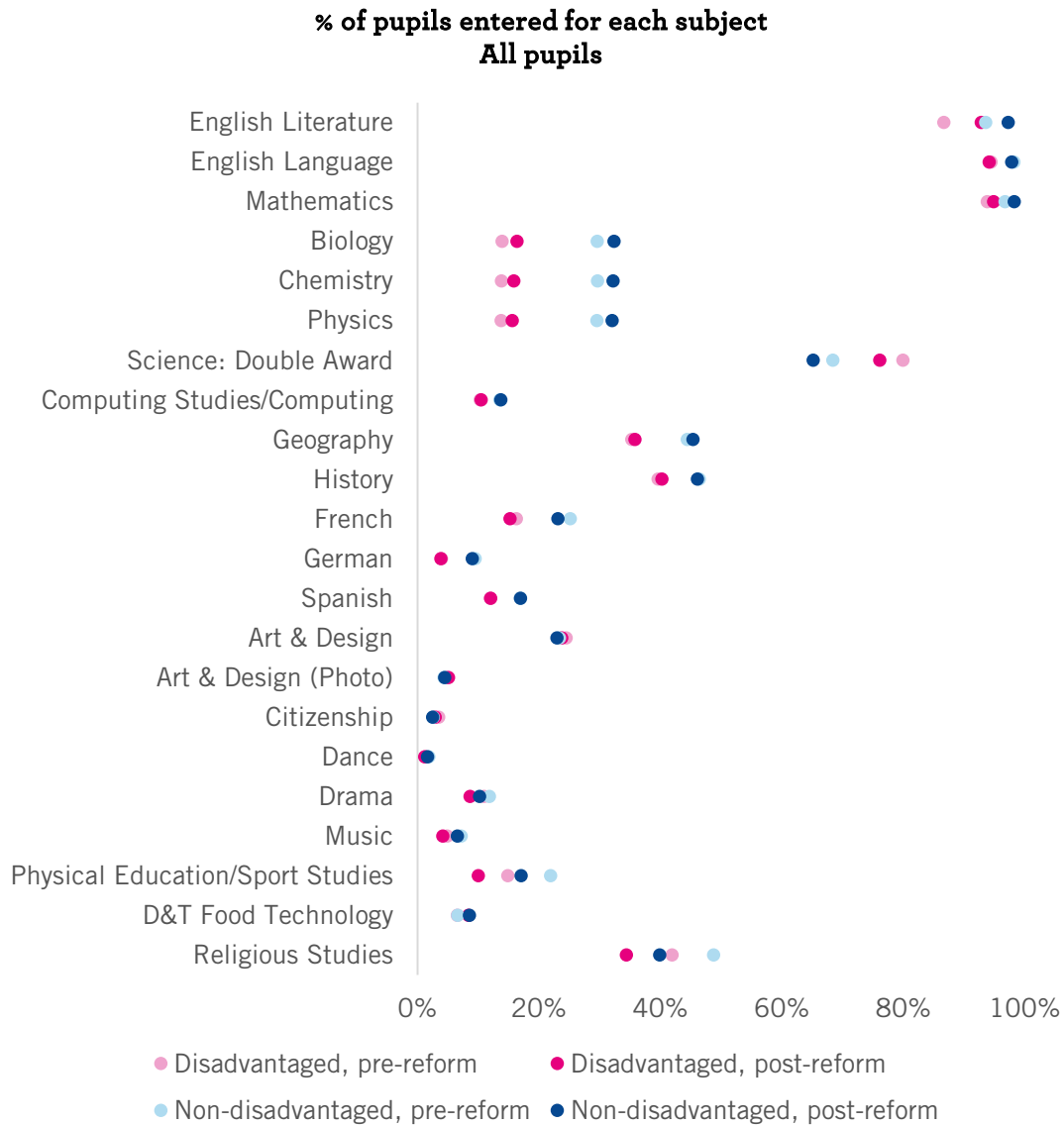
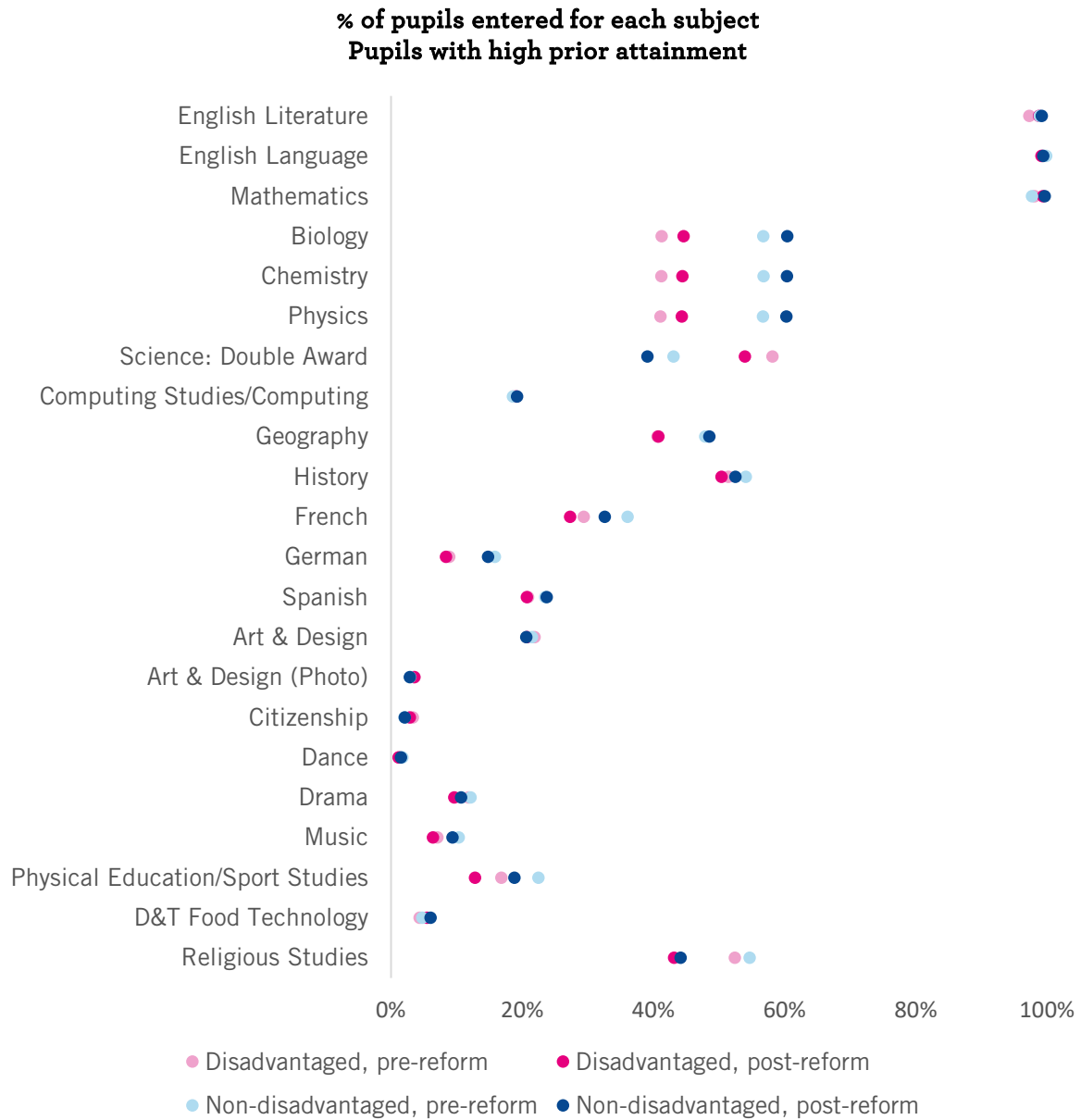


Figure 2: Percentage of pupils entered for each subject by disadvantage, pupils with high prior attainment in state-funded schools in England



Figures 1 and 2 suggest a slight drop in entry rates for both disadvantaged and non-disadvantaged pupils in most subjects. In terms of overall entries, pupils are now taking slightly fewer GCSEs in reformed subjects than they were before the reform (Table 4). The gap in entries between disadvantaged pupils and other pupils has narrowed slightly as most of the decline has been among non-disadvantaged pupils. This could be due to various reasons, including schools increasing the amount of curriculum time to English and maths or switching to non-GCSE alternatives in some subjects (for example from GCSE PE to BTEC sport studies).

Table 4: Mean number of entries in GCSEs which had reformed by 2018

	2016	2017	2018
Disadvantaged pupils	7.8	7.9	7.7
Non-disadvantaged pupils	9.0	8.9	8.7
All pupils	8.7	8.6	8.4
Entry gap	-1.2	-1.0	-1.0

Summary data on attainment in each subject for each year from 2016 to 2018 is contained in the Appendix. Table A3 shows the percentage of all pupils achieving grades 9-4 (A*-C) and Table A4 shows the same information for pupils with higher levels of prior attainment. Table A5 shows the percentage of all pupils achieving grades 9-7 (A*-A) and Table A6 shows the same information for pupils with higher levels of prior attainment.

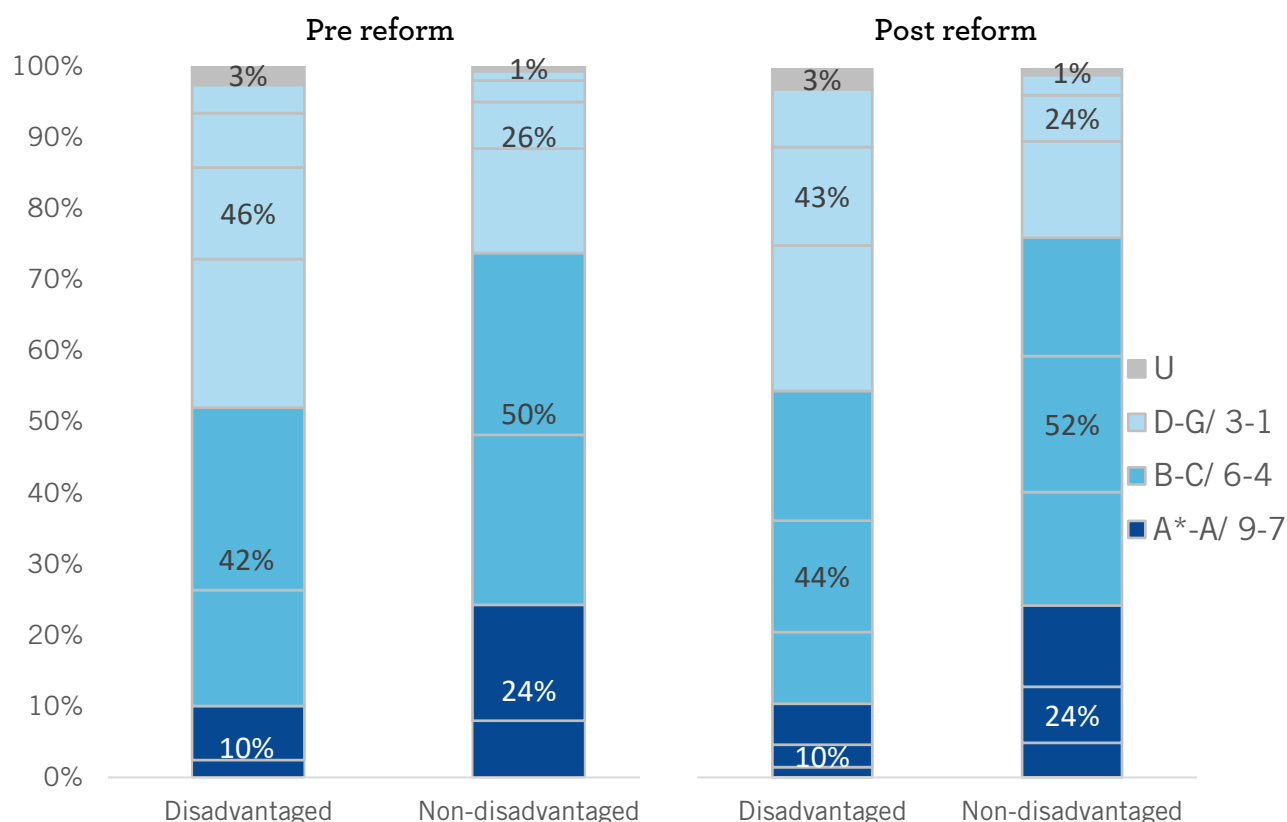
GCSE reforms and the attainment gap

We report our main attainment results, establishing the overall pupil-level change in the GCSE attainment gap from the reform. Second, we examine the attainment gap at the two grade levels seen as key in the pre-reform period. Third, we report subject specific changes in the attainment gap. We begin with some simple charts, and then summarise the overall outcomes with more formal regression analysis.

The comparison between pre-reform and post-reform is based on the following proposed equivalencies between the old and new grades. The range 9-4 in reformed GCSEs is equivalent to the range A*-C in legacy GCSEs. Similarly, the range 9-7 is equivalent to the range A*-A. These are considered the two most important grade boundaries.

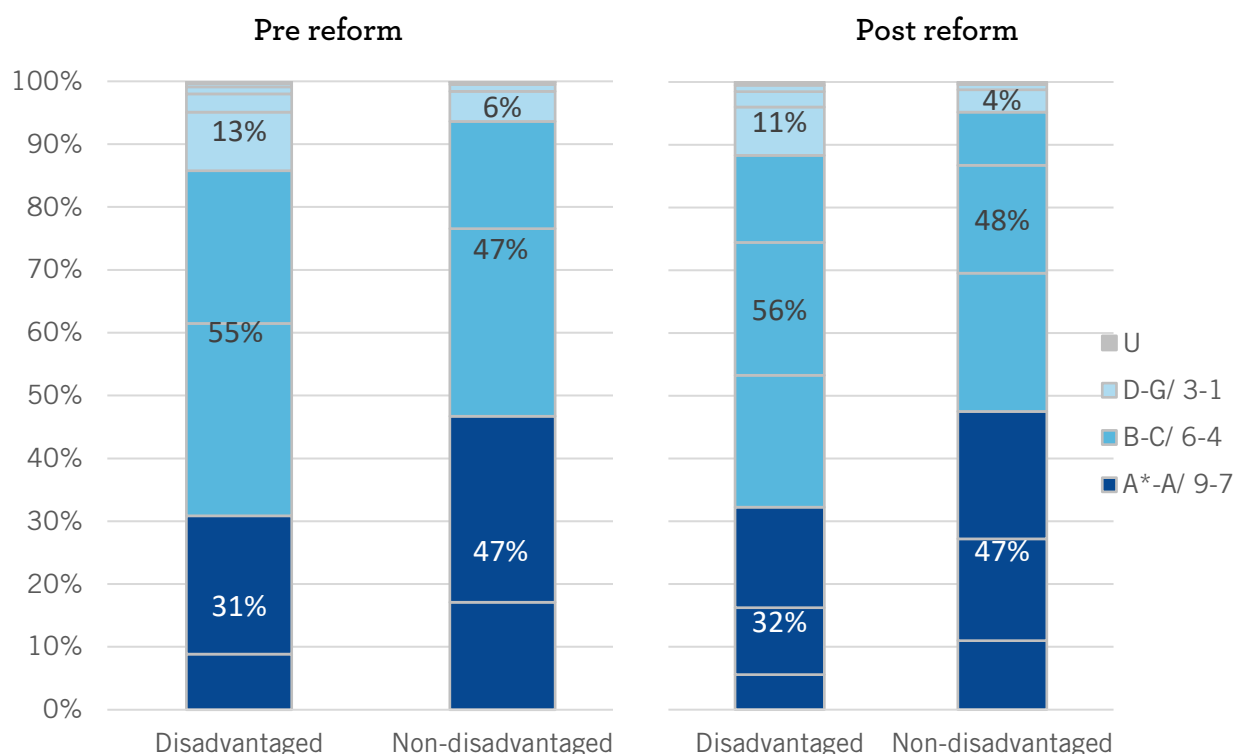
Figure 3 shows a slight increase in attainment at grades 9-4 (A*-C) following the reform for both disadvantaged and non-disadvantaged pupils. Prior to the reforms, 52% of entries in reformed subjects made by disadvantaged pupils were graded A*-C. Among the non-disadvantaged pupils the equivalent figure was 74%. Following the reforms, the percentage of entries graded 9-4 improved by 2.5 and 2.4 percentage points respectively. There is little perceptible difference at grades 9-7 (A*-A). 10% of disadvantaged pupils' entries were at these grades both prior to and following the reforms. Similarly, the attainment rate was unchanged for non-disadvantaged pupils at 24%.

Figure 3: Attainment at each grade pre- and post-reform, all pupils in all subjects, state-funded schools



Results for pupils with high prior attainment are similar (Figure 4): Attainment post-reform is broadly similar to attainment prior to the reform.

Figure 4: Attainment at each grade pre- and post-reform, pupils with high prior attainment in all subjects, state-funded schools



In order to summarise all that information, we turn to a statistical model.

Main pupil-level results

Approach

Our main approach uses a regression model to focus in on the central question: did the reforms increase or reduce the GCSE attainment gap? We account for all fixed school factors (observed and unobserved) such as the effectiveness of teachers, school funding and so on. We also account for the prior attainment of the pupils, for other pupil characteristics, and for subject differences. The analysis covers all GCSE subjects that were reformed (including predecessor subjects in English, science and food & nutrition).

We convert pupils' grades in GCSE subjects to normalised scores with mean 0 and unit standard deviation. This is carried out separately for each subject for each year. Because the reform we are evaluating is a change in the grading itself, we have to adopt a slightly more complex approach to normalisation. The cumulative percentage of pupils at each grade is tabulated from high to low, and the corresponding percentile is converted to a normalised score using the inverse normal distribution.

The central factors in the model are whether a pupil is disadvantaged, and an indicator of when the subject GCSE was reformed. The main interest is the interaction between these two: what happened to scores for disadvantaged pupils, relative to their more affluent peers, after the reform? We are also interested in whether this effect varies with prior attainment.

Results

Summary results are presented in Table 5 and full results in appendix Table A1. We first report a base model with subject and school fixed effects. This yields a value for the parameter of interest of -0.015. That is to say, relative to non-disadvantaged pupils, test scores fell for disadvantaged pupils by 1.5% of a standard deviation *per subject* after the reforms. This is a statistically significant effect and, while small, is not trivial.

Table 5: Main results

Dependent variable: Standardised GCSE score (standard errors in parentheses)

	Base model	Adding prior attainment	Adding pupil characteristics ¹⁸	Effect varies by prior attainment
FSM6*Reformed	-0.015 (0.003)	-0.011 (0.003)	-0.011 (0.003)	- -
FSM6	-0.378 (0.003)	-0.233 (0.002)	-0.236 (0.002)	-0.236 (0.002)
Reformed	-0.001 (0.002)	0.01 (0.002)	0.007 (0.002)	0.007 (0.002)
FSM6*Reformed*Low KS2				0.024 (0.004)
FSM6*Reformed*Mid KS2				-0.024 (0.003)
FSM6*Reformed*High KS2				0.002 (0.004)
Subject dummies?	Y	Y	Y	Y
Prior attainment?		Y	Y	Y
Pupil characteristics?			Y	Y
Number of entries (thousands)	12,184	12,184	12,184	12,184
Number of pupils	1,558	1,558	1,558	1,558
Measure of fit R ²	0.17	0.36	0.39	0.39

Note: Standard errors are clustered by school

The size of the effect reduces slightly as a result of controlling for prior attainment and other pupil characteristics (age, gender, ethnicity and first language).

Finally, we address whether this effect varies with prior attainment, by interacting the interaction between reform and disadvantage with Key Stage 2 bands. Disadvantaged pupils with middle levels of prior attainment appear to be the most affected, achieving -0.02 of a standard deviation lower following the reform. By contrast, disadvantaged pupils with lower levels of prior attainment benefit slightly from the reform. There is no effect of the reform on disadvantaged pupils with higher levels of prior attainment.

¹⁸ Pupil characteristics are: gender, month of birth, ethnicity, first language

We should be clear that most of these effects are small margins. And they are identified from a simple interaction of time/cohort and disadvantage. We would expect some natural fluctuation between cohorts if they differ in composition, and we would not want to ascribe this natural change to the policy. These results showing marginal changes should be treated with appropriate caution.

Attainment at key grades

Although the reformed GCSEs were designed to be more rigorous, comparability with legacy GCSEs was also a requirement. The range 9-4 in reformed GCSEs is equivalent to the range A*-C in legacy GCSEs. Similarly, the range 9-7 is equivalent to the range A*-A. To investigate the effect of the reform on these key grade boundaries, we first simply rerun the main regression above,¹⁹ but focussing on two binary variables in turn: whether the student scored at least a C/at least a 4; and whether the student scored at least an A/at least a 7. Each analysis again pools all subjects, along with all the pupil characteristics but without school-level fixed effects.²⁰

The models are fitted using logistic regression and Table 6 presents the resulting odds ratios. Full results can be found in Appendix Table A2.

Table 6: Impact of reforms on probability of achieving key grades (odds ratios)

Standard errors in parentheses

	Probability of at least grade C/4		Probability of at least grade A /7	
	Base model	Effect varies by prior attainment	Base model	Effect varies by prior attainment
FSM6*Reformed	0.959 (0.006)	0.924 (0.007)	1.002 (0.009)	1.010 (0.013)
FSM6	0.404 (0.004)	0.497 (0.004)	0.382 (0.006)	0.536 (0.007)
Reformed	1.072 (0.006)	1.131 (0.008)	1.013 (0.006)	0.986 (0.008)
FSM6*Reformed*Low KS2		1.018 (0.014)		1.005 (0.044)
FSM6*Reformed*High KS2		1.003 (0.017)		1.007 (0.016)
Subject dummies?	Y	Y	Y	Y
Prior attainment?		Y		Y
Pupil characteristics?		Y		Y
Number of entries (thousands)	12,184	12,184	12,184	12,184
Number of pupils	1,558	1,558	1,558	1,558
Measure of fit R ²	0.13	0.29	0.15	0.26

Notes: Middle KS2 attainment is used as a base category

Final model also includes FSM6*Prior attainment and reformed*prior attainment dummies and pupil characteristics¹⁸

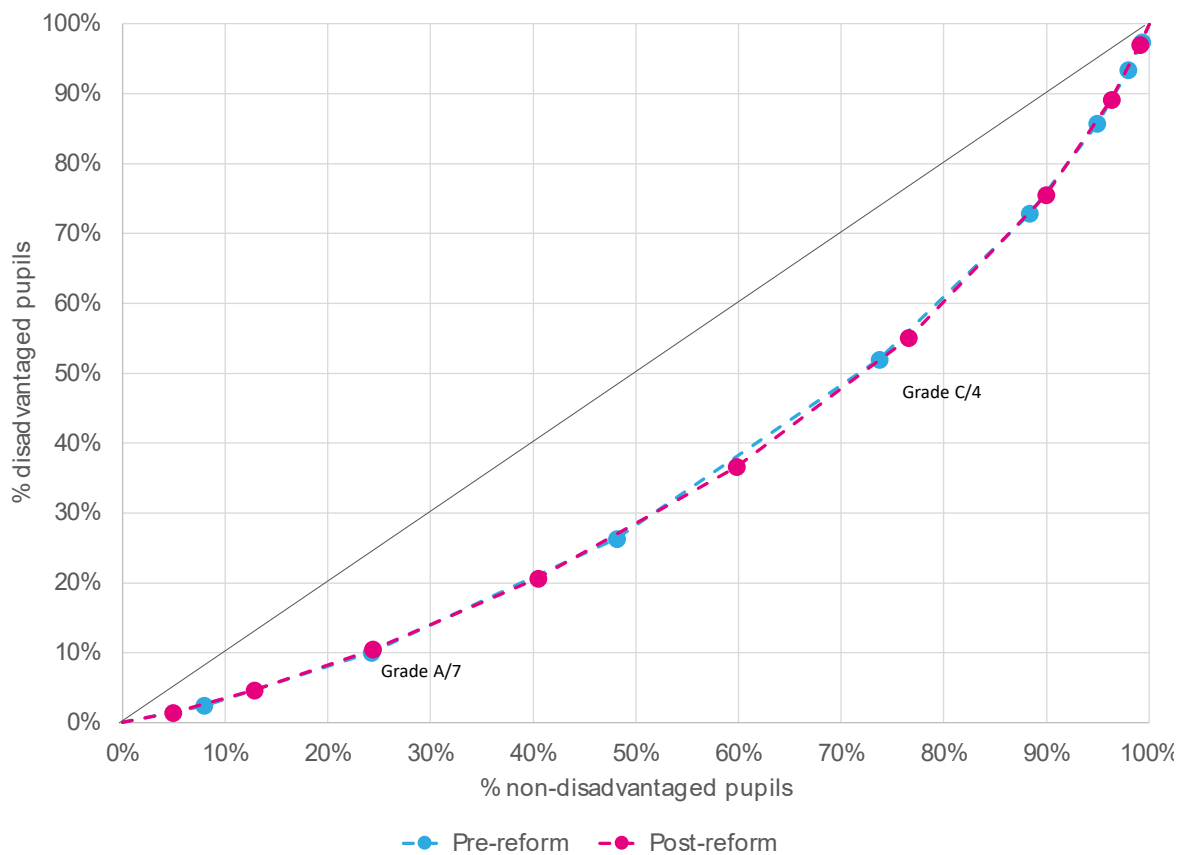
Standard errors are clustered by school

²⁰ This is because 0% or 100% of pupils at some schools achieve the threshold

These results again suggest very small differences in the relative chance of achieving these grades after the reform. At grade 7/ A, even with the extremely large number of observations, none of the interactions in Table 6 are significantly different from zero, indicating that across these grade boundaries, there was little change in the attainment gap. At grade C/4, the interaction of FSM6 and reformed status was significant- albeit small- in both models, indicating slightly lower post-reform results for disadvantaged pupils at this grade boundary once changes in subject entry patterns are taken into account. This effect was found among disadvantaged pupils with middle levels of prior attainment.

Considering the results of Tables 5 and 6 together, the fact of an overall change in the attainment gap, but little change at the ‘headline’ grade boundaries implies that there is slightly more change elsewhere in the score distribution. This is true, as we can see in Figure 5:

Figure 5: Attainment at each grade pre- and post-reform, all pupils in all GCSE subjects



This shows the percentage of non-disadvantaged students achieving at least a particular grade on the horizontal axis, against the same measure for disadvantaged students on the vertical axis. The graph plots these numbers for the pre-reform grades and the post-reform grades. The gap between each point and the 45 degree line shows the extent of the attainment gap at that grade. For example, at grade B: 48% of non-disadvantaged students achieved at least a B, compared to 26% of disadvantaged students.

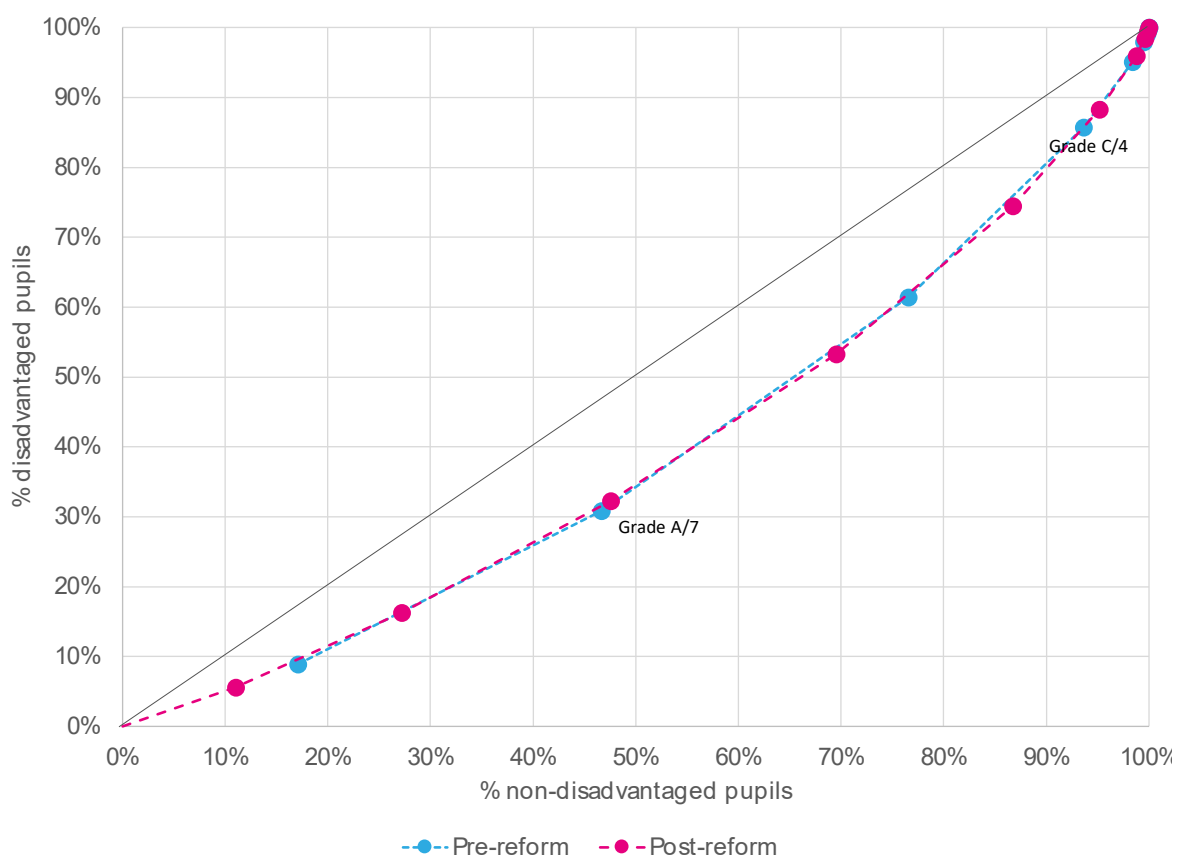
Two points are clear from the graph. First, at the key boundary of A/7, the new and the old grading systems coincide almost perfectly. This mirrors the findings in Table 6 above. There is some difference at the other key boundary of C/4, with a slightly higher percentage of both groups of pupils achieving this standard following the reform. This is likely to reflect changing entry patterns for lower attaining pupils (with pupils taking fewer subjects post-reform), since the comparable outcomes approach is

designed to maintain grade standards provided there is no change in the prior attainment profile of pupils entering a particular subject.

Second, the other grade points from the two regimes do not overlap; they obviously can't as there are more grades available, more differentiation, after the reform. Nevertheless, it is clear that the overall pattern of the grades in both regimes are very similar: the lines by and large overlap each other, and the small deviations of the post-reform line below the pre-reform line, particularly at Grade 5, illustrate the small widening of the overall attainment gap show in Table 2.

The picture is broadly similar for pupils with higher levels of prior attainment at Key Stage 2. At the higher end of the attainment distribution, the post-reform lines are fractionally higher than the post-reform line. However, they are again slightly lower at the Grade 5 boundary.

Figure 6: Attainment at each grade pre- and post-reform, pupils with high prior attainment pupils in all GCSE subjects



Overall then, there is very little difference in the attainment gap at the 'headline' grade boundaries, getting at least C, and getting at least A. However, if there is a slight drift in focus so that other boundaries take on greater importance, then that may change. For example, if as noted above the headlines became the percentage getting grade 5 or above, and the percentage getting grade 9, there would be a different story as Table 7 shows.

Table 7: Attainment pre-reform and post-reform by disadvantage, all subjects

	Expected standards		Highest grades	
	% at least C (pre-reform)	% at least 5 (post-reform)	% grade A*	% 9
Disadvantaged pupils (FSM6)	52%	37%	2%	1%
Non-disadvantaged pupils	74%	60%	8%	5%

Non-disadvantaged pupils are 3.37 times more likely to achieve grade 9 than disadvantaged pupils. Prior to the reforms, they were 3.26 times more likely to achieve grade 8. Similarly, they were 1.63 times more likely to achieve grade 5 or higher following the reform whereas they were 1.42 times more likely to achieve grade C or higher beforehand.

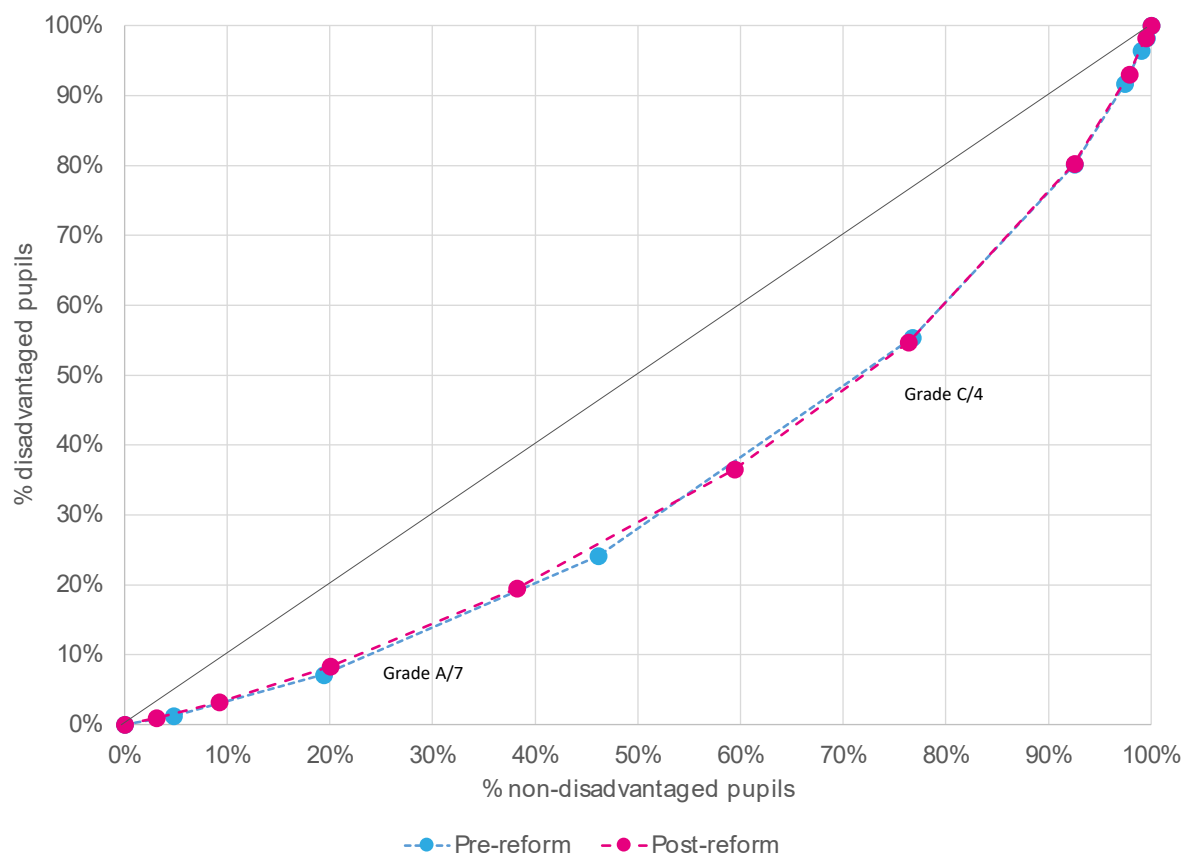
Changes by subject

Attainment in English and mathematics

Attainment in English and mathematics has long been seen as critical for future education and labour market prospects. Since 2015, young people who do not achieve at least a grade 4 (previously grade C) must continue to study these subjects as part of their post-16 study programmes.²¹

Figure 7 shows the attainment gaps at each grade between disadvantaged pupils and their peers in GCSE English before and after the reform. At the key grade boundary of C/4, the points almost perfectly coincide, indicating little change in attainment for both groups. Elsewhere in the distribution, the post-reform line is plotted below the pre-reform line at grade 5 but there appears to be a slight closing of the gap at the higher grades.

Figure 7: Attainment at each grade pre- and post-reform, all pupils in GCSE English language

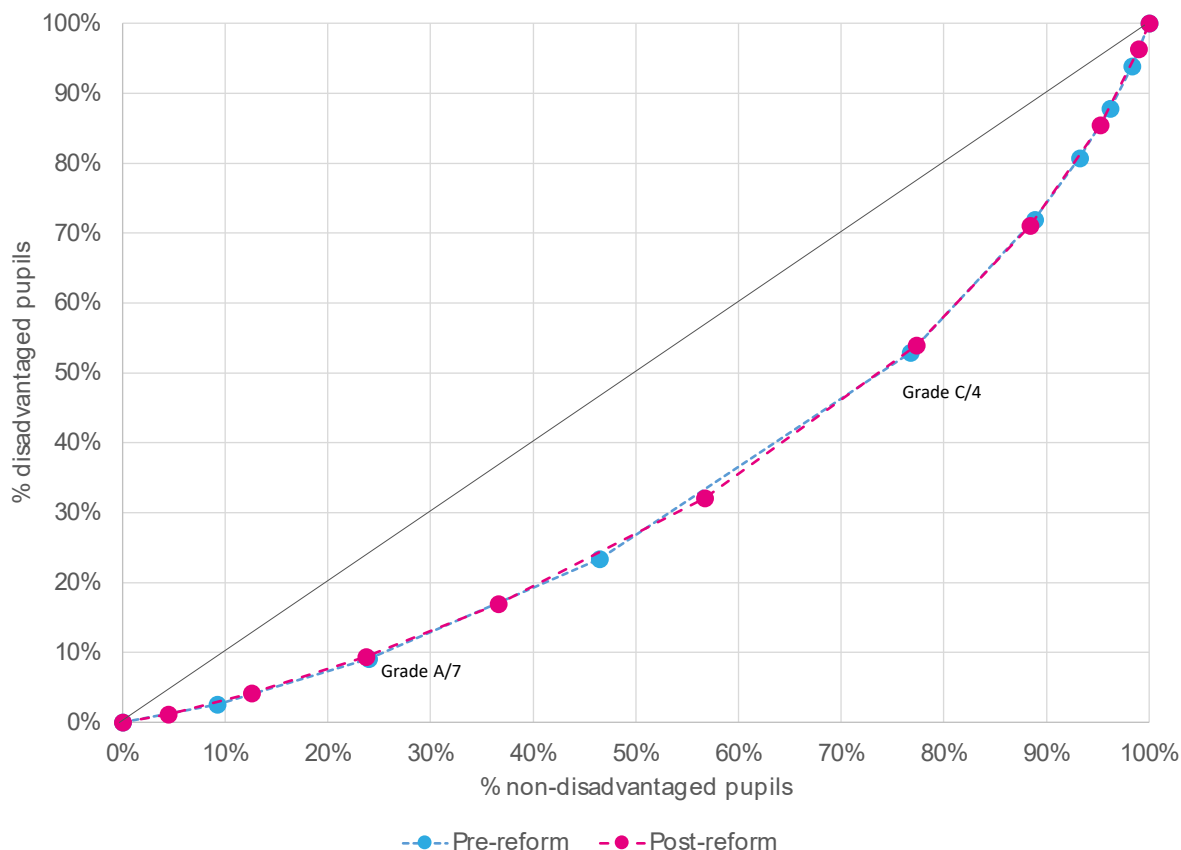


In mathematics, Figure 8 shows little change in the attainment gap following the reform.

²¹

https://www.google.com/search?q=post+16+conditions+of+funding&rlz=1C1CHBF_enGB819GB819&oq=post&aqs=chrome.0.69i59j69i57j0j69i60i3.790j0j7&sourceid=chrome&ie=UTF-8

Figure 8: Attainment at each grade pre- and post-reform, all pupils in GCSE mathematics



Pupils with a highest grade of 3 in either GCSE English language or GCSE English literature must continue to study GCSE English language post-16. Similarly, pupils with grade 3 in GCSE maths must continue to study GCSE maths. Pupils with lower grades can pursue ‘stepping stone’ qualifications in literacy and numeracy instead of GCSEs.

In Table 8, we show the percentage of the full cohort of pupils, including those without any GCSE entries, who did not achieve at least a grade 4/C. Since the reforms to GCSE, proportionately fewer pupils have been required to retake GCSEs because slightly fewer pupils achieve grade 3 than achieved grade D previously.

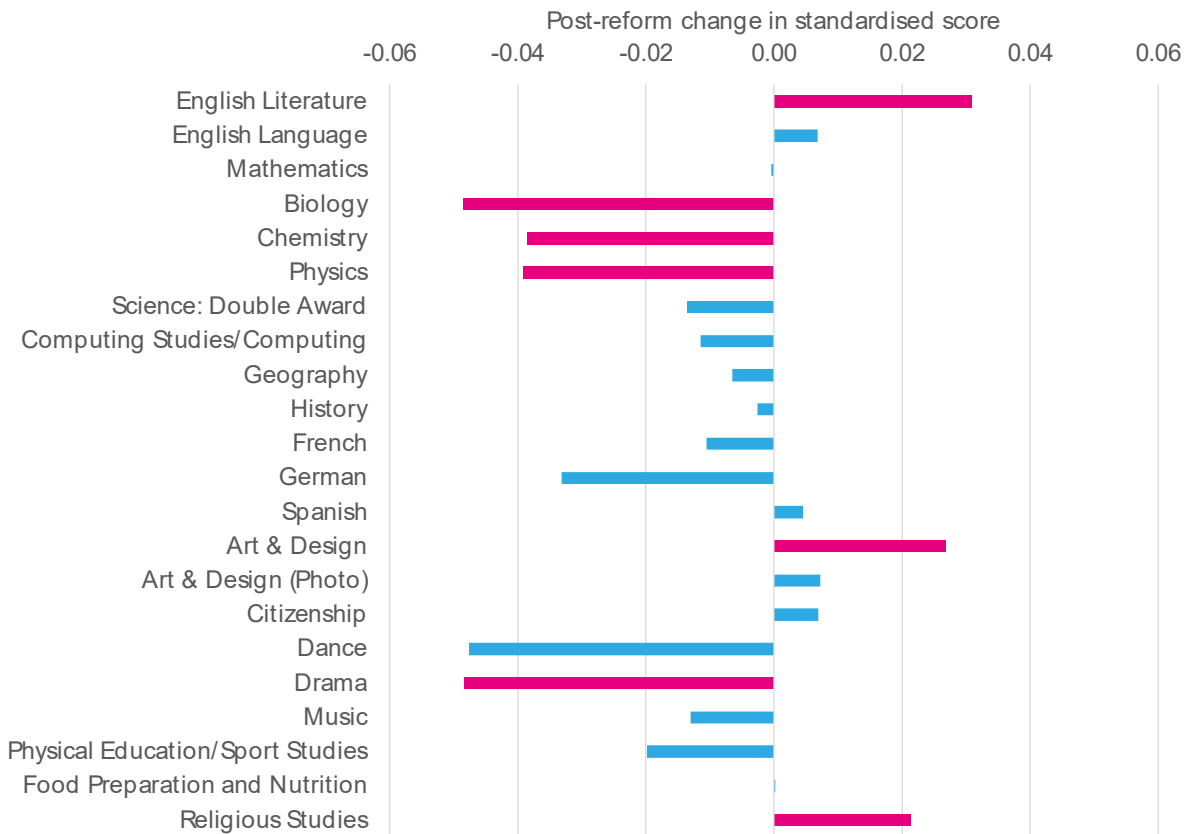
Table 8: Pupils below grade 4/C in English and maths, 2016-2018

		% Below grade 4/C		% Grade 3/D	
		All pupils	Disadvantaged pupils	All pupils	Disadvantaged pupils
English	2016	25.2%	41.1%	13.8%	19.6%
	2017	24.4%	39.9%	12.4%	18.0%
	2018	24.5%	40.1%	13.0%	18.6%
Maths	2016	31.3%	49.8%	13.5%	18.0%
	2017	30.7%	49.0%	12.1%	15.8%
	2018	30.5%	48.8%	12.5%	16.6%

Variation between subjects

The overall results presented above consider all subjects together. However, there are variations between subjects. To show this, we re-run the base model described above for each subject individually (i.e. removing the subject dummies and instead running the model separately for each subject). Below in Figure 9 we plot the key coefficient for each subject: the value of the interaction of disadvantage and reform, with figures below zero indicating a negative relative impact in disadvantaged pupils in that subject, and figures above zero indicating a positive impact. Puce coloured bars indicate results which are statistically significant.

Figure 9: Variation in the Post-reform Attainment gap by Subject



This shows that the effect on Maths was essentially nil, and on English Language slightly positive. But results in the triple sciences are negative. For the remaining optional subjects, there is variation but the majority see negative effects. That is: for most but not all subjects, including central ones like science, the impact of the reform tended to slightly raise the GCSE score attainment gap. The only strong pattern we see in grouping subjects is in the science GCSEs, widening the test score gap.

Summary and conclusions

The central aims of the reforms of GCSEs were greater differentiation of pupils and more challenging content. These ideas were encapsulated by the call for greater rigour in the testing. The reforms can be evaluated on a number of outcomes, but the one pursued here is to gauge the impact on the GCSE score gap between disadvantaged and non-disadvantaged pupils. We have explored the early impacts on attainment and on the subjects chosen by pupils.

Our central finding is that the reform has increased the GCSE test score gap between disadvantaged and non-disadvantaged pupils. The change is small, at an average of 0.02 standard deviations per subject, but is not trivial and is statistically significant. The results showed that the worsening of the gap is found among pupils with middle levels of prior attainment- those who were at level 4 on average in Key Stage 2 reading and maths tests. The effect of the reform on disadvantaged pupils with higher levels of prior attainment was neutral.

As we noted, GCSE scores are the key gatekeeper to strong life chances. This worsening of the attainment gap might therefore be expected to have implications for educational inequality and social mobility. Previewing the reforms in 2012, one of us wrote:²²

“It seems inevitable that one implication of this [the GCSE reforms] (in the short run at least) is greater educational inequality. One outcome of making an exam more rigorous is that more pupils will fail it. It seems hard to see how that could not be true. This is not fatalism; it may well be that over time, teaching brings more pupils above the line, but there will be a number of cohorts first with more failures. A tougher exam will not be credible if more people pass it.”

In fact, this has not really come to pass: while the rates of all pupils achieving the highest grades fell (by design), there has been little change in the overall pattern at the lower end of the distribution. The point remains that an exam system made “much more rigorous” might be expected to generate more failing grades as well as slightly fewer top grades. It seems likely that the operation of “comparable outcomes” and very careful matching across grade boundaries meant that the overall distribution of marks was not dramatically different.

The impact specifically on the failure rate has been somewhat masked by linguistic innovation around ‘failing’. The fraction of pupils failing – in the sense of not getting a “standard pass” – has remained the same, by design. In terms of grades and current words, the desired push for harder exams and greater challenge has been entirely and precisely offset by setting the grade boundaries to harmonise the failure rate.

However, also introduced was the descriptor “strong pass” (defined by grades 5 and above) in contrast to the “standard pass” (4 and above). This of course implies a much higher “failure” rate, and so more in line with delivering greater rigour.

We have shown that there is very little difference in the attainment gap at the ‘headline’ grade boundaries, getting at least C, and getting at least A. If the new version of these particular markers in the distribution remain the ones of central interest, then the impact of the reform on this outcome will be relatively neutral. But if there is a slight drift in focus so that other boundaries become the salient splits talked about, then that may change.

²² <https://cmpo.wordpress.com/2012/09/27/changing-the-exam-system/>

It is at the grade 5 boundary where most of the negative effect of the reform on disadvantaged pupils occurs. Non-disadvantaged pupils were 1.63 times more likely to achieve grade 5 or higher following the reform whereas they were 1.42 times more likely to achieve grade C or higher beforehand. This will matter if Grade 5 rather than Grade 4 becomes the expected standard for progression to post-16 courses or even in university admission.²³

The new grading system has – by design – delivered greater differentiation between pupils. If the grades you get matter for study and job success, then this greater differentiation will lead to greater inequality.²⁴ Not by design but perhaps inevitably, the reforms have also slightly increased the test score gap between poor pupils and the rest, which is also bound to increase inequality. These are both relatively small changes, but they both go in the same direction of raising inequality.

These are short run results, just one or two years after the changes. Two things are important for considering whether the longer term might be different. First, both teachers and pupils might adapt to the more challenging curriculum, and scores could improve. The nature of any further changes in the GCSE attainment gap depends on who adapts the best: disadvantaged or non-disadvantaged pupils; and the teachers of disadvantaged pupils or the teachers of non- disadvantaged pupils.

The second long-run issue revolves around the debate on what education actually does. If it is simply a sorting mechanism (accurately labelling the high- and low- ability pupils), then the fact that the grading offset the curriculum change is important and will mean little change in true skills, or labour market outcomes. If education does causally raise human capital, as the evidence suggests, then the curriculum change will matter and will change the distribution of skills and labour market outcomes; in this case the grading offset and the greater differentiation may matter less.

In either case, it is not clear that these mechanisms will do anything to ameliorate the worsening attainment gap at GCSE due to the 2015 reforms.

Appendix tables available from: <https://www.suttontrust.com/research-paper/making-the-grade/>

²³ <https://www.bbc.co.uk/news/education-40418457>

²⁴ Absent some strong, strange and counter-intuitive behavioural responses.

The Sutton Trust
9th Floor
Millbank Tower
21-24 Millbank
London, SW1P 4QP

T: 020 7802 1660

F: 020 7802 1661

W: www.suttontrust.com

Twitter: [@suttontrust](https://twitter.com/suttontrust)

