WHAT WORKS AT KEY STAGE 4 IN TERMS OF IMPROVING GCSE OUTCOMES, TWO OR THREE YEARS OF STUDY?

Evaluation Report

January 2023

Helen Poet, Chris Morton, Simon Rutt, Afrah Dirie, Megan Lucas and Matthew Walker

National Foundation for Educational Research

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education. Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:

Jonathan Kay
Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

This project was independently conducted by a team from the National Foundation for Educational Research (NFER). The NFER is the leading independent provider of education research and holds the status of Independent Research Organisation (IRO) from UK Research and Innovation (UKRI). Our unique position and approach delivers evidence-based insights designed to enable education policy makers and practitioners to take action to improve outcomes for children and young people. Our key topic areas are: accountability, assessment, classroom practice, education to employment, social mobility, school funding, school workforce and systems and structures. As a not-for-profit organisation, we re-invest any surplus funds into self-funded research and development to further contribute to the science and knowledge of education research **www.nfer.ac.uk@TheNFER**.

Contact details:
Simon Rutt, Head of Statistics
National Foundation for Educational Research (NFER)
The Mere, Upton Park, Slough SL1 2DQ
Email: enquiries@nfer.ac.uk
Tel: 01753 574123

Acknowledgements

The research team would foremost like to thank the school staff that participated in the research by completing the survey or by participating in an interview. Without their participation this research would not have been possible.
We would also like to thank our colleagues at NFER that contributed to the research. We would like to thank Dan Finn, Kathryn Hurd, Shazia Ishaq and Jishi Jose for their operational expertise in implementing the survey and cleaning the data. We would also like to thank Michelle Judkins who conducted interviews and Vrinder Atwal for her administrative support.

We greatly appreciate the input from colleagues at the Education Endowment Foundation, particularly Jennifer Stevenson, and the input from the Advisory Group: Jake Anders and Kim Bohling.

Executive Summary

The project

This study was funded through the Education Endowment Foundation's 2019 'School Choices' round. It aimed to investigate the extent of, and rationale for, offering a two- or three-year Key Stage 4 (KS) in England and explore how the choices made by schools about their length of KS4 affected pupil outcomes at GCSE. This was one of the EEF's first-commissioned School Choices studies, which consider whether and how the different choices schools make lead to different outcomes. These studies examine variation in practice within the education system, rather than evaluating a specific intervention, and typically use quasi-experimental designs (QEDs) to estimate the impact of the different approaches.

The structure of the national curriculum for secondary schools in England is that KS3 runs from Year 7 to Year 9 and KS4 from Year 10 to Year 11. While the national curriculum is compulsory for maintained schools, academy schools (around three-quarters of secondary schools in England) may choose whether or not to follow it and consequently some schools have opted to change their curriculum, including its length.

As is typical for School Choices studies, this study was a QED. It used a matched difference-in-differences approach to investigate the differences in outcomes at GCSE for pupils in schools offering a three-year KS4 compared to a two-year KS4. The primary outcome was GCSE mathematics, with GCSE English literature, five A* to C GCSEs, and curriculum breadth as secondary outcomes. Schools from the three-year KS4 arm (those starting KS4 maths in Year 9 rather than Year 10) were matched to schools from the two-year KS4 arm that were similar in terms of their observable baseline characteristics using a propensity score approach.

A total of 405 schools responded to a survey exploring KS3 and KS4 length for their Year 9, 10, and 11 students. The impact analysis included 104 of these schools (170,675 pupils); the data was matched to the National Pupil Database for analysis. This group includes schools that first converted to a three-year KS4 for maths for exam years between 2014 and 2018 and their matched two-year schools. The study included all pupils who were present at the start of the KS4 programme from participating schools in the academic years 2007/2008 to 2018/2019.

The implementation and process evaluation (IPE), which drew on evidence from a survey and telephone interviews with school leaders, focused on the reasons for offering different models, perceived impacts, views on curriculum breadth, and the influence of Covid-19 on the curriculum.

The study ran from September 2019 to December 2022.

Table 1: Key conclusions

Key conclusions
The research is unable to conclude that any differences in observed outcomes are due to the length of KS4. This is because it was not possible to achieve a strong match through the QED, with the two groups of schools on different GCSE performance trajectories prior to the KS4 length policy change. This caveat applies to the primary outcome (maths attainment), the secondary outcomes (English literature and 5 A*-C grades at GCSE), and to a subgroup analysis on the maths performance of everFSM pupils[1]. The independent evaluation team does not interpret any differences observed as causal - differences in outcomes may have been caused by other factors alternative to, or in addition to, changing the length of KS4.
The evaluator's recommendation is that schools should not make a decision about the length of their KS4 on the basis of the impact evaluation (QED) findings reported here.
Almost twice as many schools responding to the survey delivered KS4 over three years (for at least some subjects) rather than over two years.
Both shorter- and longer-length KS4 schools were offering - and pupils were taking - fewer qualifications at the end of KS4; the number of qualifications declined from a peak in 2011/2012 to the time of the survey in 2019/20. Analysis of curriculum breadth (a secondary outcome measure), and the IPE findings, indicated that this shift was driven by policy changes relating to school-level performance measures, such as the EBacc, and changes to the way that vocational qualifications contribute to school tables.Curriculum breadth and depth was a key consideration for schools irrespective of their curriculum model.

[1] 'EverFSM pupils' are those that have ever been eligible for free school meals (FSM).

Schools running a three-year KS4 were motivated to do so by their views of the requirements of the new GCSEs and, to a lesser extent, to improve pupil engagement in Year 9. In contrast, schools that had maintained a two-year KS4 described their primary motivation as the importance of a strong curriculum and breadth of experience at KS3 when delivered over three years. Regardless of the length of their KS4, schools emphasised the importance of tailoring the curriculum to their specific intake and context to best support their pupils. Schools often considered KS3 and KS4 holistically and reviewed the sequencing and delivery of the curriculum across the five years, rather than as two distinct key stages.

EEF security rating

These findings have a low to moderate security rating (two padlocks). This was a quasi-experimental study that used a matched comparison group, difference-in-differences approach. There were important differences in almost all observable characteristics at baseline of schools running a two-year KS4 compared to those with a three-year KS4 and there was evidence that schools were on different trajectories in terms of their GCSE performance prior to the change in KS4 length. The authors identify that many schools made the change to a three-year KS4 programme as part of a wider suite of other policy changes and are unable to account for this in the analysis. This means that it is not possible to disentangle the impact of the respective length of KS4 from other policy changes that were taking place in schools at the same time.

Additional findings and associated caveats

The results of this study should not be interpreted as conclusive evidence of the effect of KS4 length on GCSE performance. Although the evaluators have estimated differences in outcomes, the findings must be interpreted with the caveat that the schools being compared were on different GCSE performance trajectories prior to the KS4 length policy change and therefore that these estimates should not be considered causal.

The study had a number of pre-specified impact analyses, and also undertook some post-hoc exploratory analysis in relation to the primary maths outcome. On the pre-specified analysis for differences in GCSE maths (the primary outcome), pupils in schools with a three-year KS4 appeared to make, on average, the equivalent of one additional month of progress compared to pupils in schools with a two-year KS4. As with any study, there is some statistical uncertainty around the result; the possible difference ranges from zero to one month of additional progress. For GCSE English literature (one of the secondary outcomes), although pupils in the three-year KS4 (maths) schools appeared to improve post policy switch, they did not improve as much in English literature as the pupils in two-year KS4 schools. The difference in English improvement between three-year and two-year KS4 schools was small: using EEF's conversion table, this equated to no additional months of pupil progress in English. Additional, exploratory, post-hoc analysis was conducted on the primary outcome of maths. This analysis accounted for the differences in GCSE trajectories prior to the change in KS4 length and found there was no difference in GCSE maths scores between the two groups of schools.

The lack of confidence in the results of the primary outcome (maths) and the English literature analysis is also applicable to a subgroup analysis that was undertaken for FSM pupils and an additional secondary outcome analysis, looking at achievement of five A* to C grades at GCSE. The subgroup analysis indicated that the maths performance of everFSM pupils in schools that converted to a three-year KS4 appeared to increase over time compared to everFSM pupils in schools that kept a two-year maths KS4. Different lengths of KS4 do not appear to have been associated with a pupil being more or less likely to achieve the five A* to C measure. These findings are subject to the above caveats related to pre-conversion trajectories – they should not be interpreted as causal and may be related to factors other than KS4 length.

The evidence from interviews with school leaders also suggests that in many of the three-year KS4 schools, the policy change may have been only part of a suite of changes or policies implemented by the schools. This adds to the challenge of attributing any differences in observed outcomes to the length of KS4 specifically.

Finally, the research found that schools do not perceive a one-size-fits-all approach to KS3 and KS4 delivery and that the participating schools had tailored their curriculum – to a lesser or greater extent – with the aim of best-serving their pupils. The survey and follow-up interviews showed that schools did not view a particular model as inherently 'right' or 'wrong'; rather, they sought a 'right approach for them and their pupils', often considering KS3 and KS4 holistically, and reviewing the sequencing and delivery of the curriculum across the five years, rather than as two distinct key stages. None of the two-year KS4 schools reported plans to change the length of their curriculum but almost a fifth of three-year schools answering the survey intended to change back to a two-year KS4; indeed, by the time of the interviews some of the schools had implemented this change.

Impact*

Table 2: Summary of impact on primary outcome(s) *

Outcome	Effect interpretation/ treatment group/ comparison group	Effect size (95% confidence interval)*	Estimated months' progress*	EEF security rating	No of pupils	P Value
GCSE maths score (primary outcome)	Difference-in-differences/three-year KS4/two-year KS4	0.05 (0.03, 0.06)	1	2 padlocks	170675	< 0.001
GCSE English Literature score (secondary outcome)	Difference-in-differences/three-year KS4/two-year KS4	-0.02 (-0.04, -0.01)	0	n/a	159600	0.012

* These estimates should be interpreted with a high degree of caution and not considered causal. Note that the additional exploratory analysis that was conducted on the primary outcome to take account of the pre-trends found no difference in GCSE maths scores between the two groups of schools. The recommendation is that schools should not make a decision about the length of their KS4 on the basis of the impact evaluation findings reported here.

Introduction

Background

This project was originally commissioned as part of the Education Endowment Foundation's 2019 round of 'School Choices' research. The EEF invited proposals for research to investigate how the different choices made by schools lead to different outcomes. In particular, The EEF was keen to understand more about practices that are difficult to evaluate using a randomised controlled trial (RCT) in situations where randomisation is not possible (for example, because of school reluctance or because it would not be appropriate to randomise a particular intervention or approach). This study, which looks at the impact of different models at Key Stages (KS) 3 and 4, certainly fits this criteria as such a significant change to a school's curriculum would not be appropriate for randomisation – nor acceptable to schools.

The structure of the national curriculum for secondary schools in England (DfE, 2014) is that KS3 runs from Year 7 to Year 9 and KS4 runs from Year 10 to Year 11. There is a broader range of compulsory subjects at KS3 than at KS4, although at KS4 schools must continue to offer pupils access to at least one subject within four 'entitlement areas' (the arts, design and technology, the humanities, and modern foreign languages). During KS4 pupils study towards qualifications such as GCSEs and BTecs. While the national curriculum is compulsory for maintained schools, academy schools (around three-quarters of secondary schools in England) may choose whether or not to follow it and consequently some schools have opted to change their curriculum, including its length.

What does the evidence tell us about the impact of a three-year KS4 compared to a two-year KS4?

Previous research about the impact of different models is limited. There is some evidence available from an evaluation of the then Department for Children, Schools and Families' Two-Year Key Stage 3 project, which launched in 2003 (Noden et al., 2007). This aimed to increase the pace of learning in KS3 and open up curricular flexibility, for example, by beginning the KS4/GCSE curricula earlier through the time saved. The evaluation focused on performance in Year 7 and Year 8 (in the then QCA Optional Tests) finding that a shortened KS3 was associated with an increased pace of learning in maths.

Why might schools move to a longer KS4?

Previous research has indicated that some schools have opted to amend how they operate KS3 and KS4 in their school, and to bring forward the start of KS4 for some or all subjects to Year 9 (i.e., up to a year earlier than the curriculum framework states). The 'Key Stage 3: The Wasted Years?' report by Ofsted in 2015 found that around a third of the schools that participated were delivering KS3 over two years (and, by implication, KS4 over three years). More recent evidence suggests that a longer KS4 has become more common, for example:

- 56% of respondents in NFER's February 2019 Teacher Voice representative omnibus survey (NFER, 2019) started teaching the GCSE curriculum for most/all subjects in Year 9 compared to 40% that started in Year 10.
- Similarly, the TES (Roberts, 2019) reported that half of ASCL members' schools offered a longer KS4.
- Ofsted's curriculum research (Ofsted and Spielman, 2017) found 10 of 23 secondary schools visited in 2017 were reducing KS3 to two years and around a quarter of 171 school websites reviewed indicated pupils were selecting GCSE options at the end of Year 8.
- DfE's School Snapshot Survey Winter 2017 (IFF Research, 2018) found that of the 282 responding schools with Year 9 pupils, 35% started KS4 in Year 9 for all subjects and 28% began KS4 in Year 9 in some subjects for all pupils. In 27% of schools, KS4 started in Year 10.

In addition to academisation, another relevant policy change was the introduction of the EBacc (first applied in the 2010 school performance tables) combined with the new GCSEs (first teaching 2015 to 2017; first testing from 2017 to 2019), which accelerated the adoption of a longer KS4 in order to provide sufficient teaching time. The change to GCSEs was outlined in a letter to Ofqual by the then Secretary for Education in 2013 explaining the requirement for GCSEs to be

'comprehensively reformed' including a move to linear qualifications with all assessments at the end of the course.[2] The new GCSEs would be designed to 'test the depth and breadth of pupils knowledge and abilities' and a new grading scale was suggested (and later adopted in the move from A*–G to 1–9 grading for some subjects from 2017). The new qualifications needed to 'evidence pupils' achievement against demanding and fulfilling content … [and] provide a strong foundation for further academic and vocational study'. They also would form the 'basis upon which schools will be held accountable for all of their pupils', reported in performance tables.

In parallel, the recommendations of the Review of Vocational Education (Wolf, 2011) were implemented. Of most relevance here was the first recommendation, which led to an update of the (vocational) qualifications that would contribute to the school performance indicators at KS4 – and those that no longer counted. This also links to the debate around the breadth and depth of the curriculum (sometimes discussed in terms of 'curriculum narrowing'), which is outlined briefly below.

Ofsted has stated that it does not have a preferred curriculum model and that inspections expect schools to offer pupils an 'ambitious' curriculum with the opportunity to study a broad range of subjects across their whole time in secondary education (Harford, 2020; Ofsted, 2020a). Ofsted's latest inspection framework (2019) focuses on both curriculum design and implementation, including recommending a broad KS3 curriculum where pupils are not expected to 'specialise too early' (Ofsted and Spielman, 2018). However, a recent blog by Middlehurst (2020) from the Schools, Students and Teachers' Network described how one of the most common questions asked by schools is about Ofsted's views on the length of KS3/KS4 and that there was a common perception in schools that 'a two-year KS3 is a limiting factor in inspection'. He went on to outline where some schools' concerns about Ofsted's views on the length of KS3 may have come from: analysis of recent inspection reports showed that 'a significant number of ["requires improvement" inspection] reports include mention of a two-year KS3 and/or a low EBacc entry rate' (Middlehurst, 2020). To counter this, Middlehurst also highlighted that at least one school with a three-year KS4 had been judged as 'outstanding' under the new framework, indicating that a three-year KS4 did not automatically lead to a low inspection rating and that 'this perception is, in fact, a myth'.

Curriculum breadth and depth

There is considerable current policy and practitioner interest and a growing body of research around the breadth of the curriculum which is relevant here. Since the Wolf review (2011), the number and equivalence of vocational qualifications eligible for the KS4 league tables has reduced and Gill (2018) noted that this appeared to have been reflected in the number and type of vocational qualifications taken. Ofsted has warned that a disproportionate or premature emphasis on teaching exam specifications limits pupils' exposure to a broad and balanced curriculum leading to curriculum narrowing, for example, through subjects being squeezed out of the KS3 curriculum (Ofsted, 2020a).

Recent NFER evidence (NFER, 2019) found that the vast majority (81%) of respondents agreed that their school offered a broad and balanced KS3 curriculum, with senior leaders answering more favourably. The majority of respondents also thought that their school's KS4 curriculum was broad and balanced, although at slightly lower levels (68%). However, 22% of classroom teachers disagreed that their KS4 curriculum had breadth and depth.

According to the same NFER research (2019), more than half (53%) of senior leaders and classroom teachers said that the total number of GCSEs taught at their schools had decreased over the last few years, with nearly half of senior leaders and teachers reporting that there were other subjects or qualifications not currently offered that they would like to include. When asked why these subjects were not currently offered, around a third of respondents said that this was because they did not fit with Progress 8, and almost as many cited staffing shortages, a lack of teaching expertise in that subject, or low take-up by pupils. Similarly, many schools found it difficult to accommodate a three-year KS3, the EBacc, and four options at GCSE leading schools to prioritise one over the others - a challenge exacerbated by the introduction of the 'new, larger' GCSEs (Middlehurst, 2020; Neumann, 2020).

More broadly, other research argues that reduced curriculum breadth can have a particular impact on disadvantaged and/or lower attaining pupils. For example, GL Assessment (2019) argued these pupils can be deprived of the benefits of a rich and broad curriculum, which their better off peers tend to accumulate outside of school through extra-curricular

[2] **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 [19 October 2022]**

activities. Ofsted has outlined concerns that the focus of the performance measures on academic subjects is particularly restrictive for low attaining pupils, which disproportionately affects pupils from low income backgrounds (Spielman and Ofsted, 2017). Furthermore, researchers at Kings College London found that teachers think the nature of the new knowledge-focused GCSEs is less engaging for these pupils with a greater risk of disaffection (Neumann et al., 2016).

On the other hand, the focus on EBacc subjects and Progress 8 has enabled more students – and in particular lower ability and/or disadvantaged students – to access more core curriculum subjects (such as science, geography, and history) as Gill's (2018) analysis of the impact of the accountability measures on the uptake of qualifications suggests.

Following the Covid-19 pandemic, the Department for Education (DfE, 2021) has also outlined the role of the curriculum in helping pupils to recover from the significant impacts the pandemic has had on their learning. The DfE urged schools to continue to teach a broad and balanced curriculum in all subjects and stated that pupils should continue to benefit from enrichment activities such as visits and external visitors in school. However, the DfE also recognised that uneven disruption across schools may mean some schools need to substantially modify their curriculum in order to adjust to the needs of pupils.

This research project

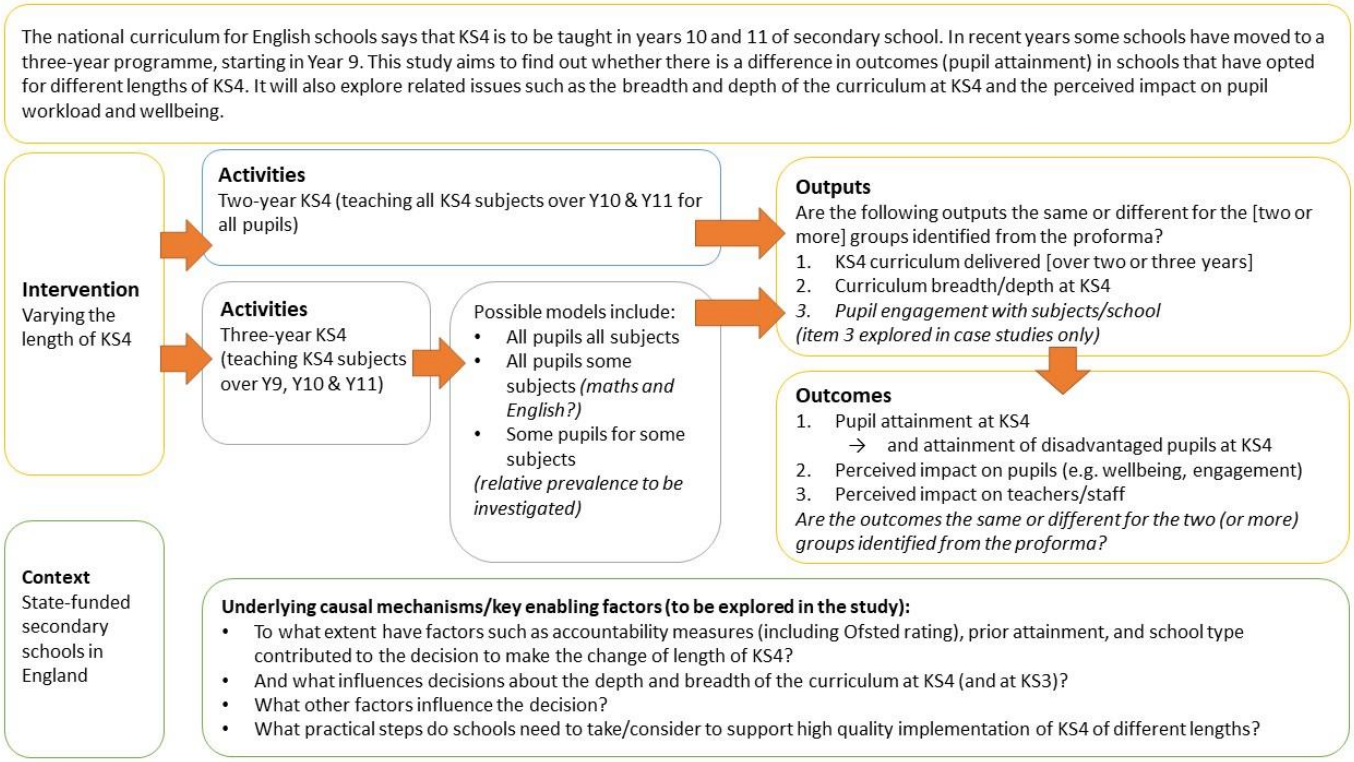
This research explores the extent and rationale of offering a two- or three-year KS4 and compares attainment outcomes at the end of KS4 in similar schools offering different lengths of KS4 delivery in some or all subjects. The research consisted of a survey sent to all secondary schools in England to find out the models used by individual schools; the responses were classified and matched to the National Pupil Database (NPD) and analysed using a difference-in-differences approach. It was supplemented by in-depth interviews with school leaders to understand the context and other factors that schools consider such as pupil wellbeing, readiness, and perceived impacts on timetabling and staffing.

Intervention: schools' choices and models at KS3 and KS4

This study explored the impact of different approaches used by schools to structure the Key Stage 3 (KS3) and Key Stage 4 (KS4) curricula. The 'intervention' is the model selected and implemented by schools. Part of the project aimed to better understand the types of provision and the prevalence of different approaches. In the study plan we identified a hypothesised logic model, which is included in Figure 1. A revised version of the logic model updated to incorporate learning from this project can be found in the Conclusions section.

Name
What works at KS4 in terms of improving outcomes, two or three years of study?
Why
Some schools have chosen to operate a three-year rather than a two-year KS4; this may be related to accountability measures (the research will explore this).
Who and where
Secondary schools in England and their Year 11 pupils.
What
The research project will look at the difference in outcomes in schools that operate a three-year KS4 compared to those schools that maintain the two-year KS4 approach.
When
The research will include KS4 results from 2007/2008 to 2018/2019.
Tailoring
The research will look at how schools have approached their curriculum at KS4 and the variations employed.

Figure 1: Hypothesised logic model



Evaluation objectives

Impact research questions

I_RQ1	Do pupils attending secondary schools that teach KS4 over three years perform differently at GCSE to similar pupils in similar schools that teach KS4 over two years? What is the impact on disadvantaged pupils?
I_RQ2	What is the impact of school-level disadvantage on attainment outcomes and its interaction with length of KS4?
I_RQ3	Do pupils in schools with two- or three-year KS4 study (on average) different numbers of GCSEs?
I_RQ4	What evidence is there that the curriculum offered by schools has narrowed, and is there a difference between schools with a two-year KS4 and those with a three-year KS4?

Implementation and process evaluation research questions

IPE_RQ1	What are the perceived impacts of having a two- or three-year programme at KS4 on pupils (in particular, but not exclusively, related to wellbeing and ability to manage workload)?
IPE_RQ2	What are the softer impacts of having a two- or three-year programme at KS4 on teachers and their lesson planning?
IPE_RQ3	What are the reasons for operating different lengths of KS3/KS4 (for example, pupil outcomes/accountability measures/resourcing factors)? What were the factors that influenced the decision to change the length of KS3/KS4?
IPE_RQ4	What form do different lengths of KS3/KS4 take? What variation occurs within groups?
IPE_RQ5	What strategies and practices are used to support high-quality implementation of different lengths of KS3/KS4?
IPE_RQ6	Which factors affect the breadth of subjects that schools offer for study at KS4 and at KS3?

The study plan can be found on the EEF project page.[3]

Ethics and trial registration

The project was subject to NFER's ethics review as part of the start-up meeting. This involved checking the project against the NFER Code of Practice checklist. The project complied with the Code of Practice and no issues requiring escalation were identified.

Schools were invited to take part in the research through the survey, which also included the opportunity for the respondent to express interest in taking part in a follow-up interview. By completing the survey, and in the case of school leaders who agreed to take part in an interview, they agreed to participate in the research.

Data protection

NFER was the data controller for this project up to the data deletion date (the EEF is the data controller for the dataset once archived).

The legal basis for processing personal data was covered by GDPR Article 6 (1) (f):

> 'Processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interest are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data.'

We carried out a legitimate interest assessment, which demonstrated that the evaluation fulfilled one of NFER's core business purposes (undertaking research, evaluation, and information activities). The project has broader societal benefits and aims to contribute to improving the lives of learners by providing evidence about the impact of the length of KS4 on pupil attainment and the perceived impacts on pupils. We therefore determined that is was in our legitimate interest to process and analyse personal data for the evaluation. We carefully considered all of the personal data being collected and all pieces were necessary to achieve the aims of the research (see Ethics, above). We also balanced any potential impact on the data subjects' rights and found that our activities would not do the data subject any unwarranted harm. Privacy information was provided to all data subjects in an accessible and transparent manner through a project-specific privacy notice and contact details for the project; our compliance officer and the ICO were made available for all participants to ask questions or raise concerns about the processing of their data.

No pupil or teacher names, nor the name of any school, are being reported. We will not share personal data collected through interviews or in the proforma or survey with other organisations.

NFER accessed the NPD data for analysis through the SRS secure online system. The SRS system does not allow users to remove or copy data from its servers.

At the end of EEF evaluations, data is archived to allow for further secondary analysis. As part of this project only school-level data will be added to the EEF archive; none of the personal data collected will be added.

[3] **https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/what-works-at-key-stage-4-two-or-three-years-of-study/**

NFER project team

Name	Roles and responsibilities
Simon Rutt	Project director, responsible for leading the NFER team and project delivery.
Helen Poet	Project manager, responsible for overseeing the day to day running of the project
Afrah Dirie	Lead statistician, responsible for statistical analysis.
Chris Morton	Statistician.
Dan Finn	Test and schools administration lead, responsible for overseeing technical design, dispatch, and processing of the proforma.
Megan Lucas	Researcher with responsibility for the case studies.
Matthew Walker	Researcher, working on IPE.
Michelle Judkins (NFER Associate)	Interviewer.

Methods

Evaluation design

This was a quasi-experimental study involving 104 mainstream, publicly funded secondary schools in England that responded to a survey regarding their approach to the structure of KS3 and KS4. It investigated the impact of the length of KS4 programme on GCSE maths attainment (primary outcome) as well as GCSE English literature and the likelihood of achieving five or more A* to C grades, and curriculum breadth (secondary outcomes). The two lengths of KS4 of interest were a three-year programme (schools teaching KS4 from Year 9 for at least maths) and a two-year programme that started teaching KS4 in Year 10.[4] Schools from the three-year KS4 arm were matched to schools from the two-year arm that were similar (in terms of the observable baseline characteristics) at the time the decision to switch from two- to three-year programme at KS4 was assumed to have been made. Matches were found using a combination of nearest neighbour and exact matching without replacement. The study included all pupils who were present at the start of the KS4 programme from participating schools in the academic years 2007/2008 to 2018/2019. The project has a difference in differences design, which compares the change in average exam scores between the two-year and three-year KS4 arms after the introduction of the three-year KS4 system. This design is used to minimise bias caused by different baseline characteristics between the arms, which would be achieved by randomisation in a randomised controlled trial. A linear multilevel regression model was used for the main analysis. Additional sensitivity analyses were conducted and further analysis to explore heterogeneity over time is also reported.

[4] The two-year KS4 group is also known as ‘never-switchers’ and the three-year group is made up of schools that switched and did not switch back according to the data they gave in the survey. Schools that switched to a three-year KS4 programme and switched back were excluded from analysis.

Table 3: Evaluation design

Evaluation design, including number of arms		Matched difference-in-differences study, two arms
Unit of analysis		Pupil
Primary outcome	Variable	GCSE maths performance
	Measure (instrument, scale, source)	Maths GCSE points awarded, 1–8 (see Table 6) NPD
Secondary outcome(s)	Variable(s)	(1) GCSE English literature performance (2) Five plus A*–C grades (or equivalent) including English and maths (see Table 6) from NPD (3) Curriculum breadth
	Measure(s) (instrument, scale, source)	(1) English GCSE points awarded, 1–8, NPD (2) Achieved at least five GCSE points between 5 and 8 (see Table 6) including English and maths 0–1, NPD (3) Number of subjects offered and/or proportion of pupils studying them per school (NPD)
Baseline for primary outcome	Variable	Overall KS2 attainment in maths
	Measure (instrument, scale, source)	Total marks achieved in KS2 maths test, 0–110, NPD
Baseline for secondary outcome(s)	Variable	(1) Overall KS2 attainment in reading (2) Overall KS2 attainment in English and maths (3) No baseline measure
	Measure (instrument, scale, source)	(1) KS2 reading test mark, 0–50, NPD (2) KS2 reading and maths test mark, 0–160, NPD (3) No baseline measure

Participant (school and pupil) selection

In order to find out which KS3/KS4 models were used by schools and when they had adopted the model, we decided to ask them directly using a short survey/proforma, because this information is not routinely collected from all schools. Originally it was stated that a minimum of 200 matched schools (100 two-year and 100 three-year KS4 schools) would be needed for the impact analysis in the study plan. As the data collection was not linked to an incentive, the expected response rate was around 10%. All secondary schools in England were sent a proforma survey in February 2020. This was to enable us to create a variable indicating the length of KS4 programme for the main impact analysis of KS4 outcomes.

All secondary schools in England (N = 3,365) were sent a school-specific link to easily access the proforma (hosted in our survey software Questback). In the first mail-out, all schools were contacted by email to maximise the response from the first contact. Letters were posted to schools for which the emails did not succeed. Reminders and social media were used as appropriate, however, the survey period was paused in late March 2020 due to the national lockdown in response to the Covid-19 pandemic. The survey was reopened briefly in June 2020, and the decision was made to close the survey in September 2020.

The survey was completed by 405 secondary schools:[5] in almost all it was completed by the headteacher or a senior leader (91%) and in most cases by someone who had been at the school for at least three years (83%).

The first set of questions were 'closed' and essential for the impact analysis; all of these questions were mandatory. Half way through the survey, respondents were given the option to exit the survey early: 79% of respondents opted to continue. The second half explored the reasons for their model and as such consisted of more 'open' questions that invited the respondent to write in their answer rather than tick a box. All respondents, whether they completed the whole survey or opted to leave at the mid-point, were invited to take part in a case study (later changed to telephone or video interviews due to Covid-19 restrictions, see below). The full survey can be found in Appendix C.

While all schools were included in the descriptive analysis (see section on IPE), we only selected schools for the impact analysis that had data that covered the period we wished to analyse. Prior to seeing the results of the survey, we expected that schools would provide a range of academic years in which they made the change to delivering KS4 over three years. Our expectation was that we would be able to obtain KS4 performance data for at least a couple of years after the first cohort that began KS4 in Year 9 sat their KS4 examinations, as well as performance data from before the school made the change to a three-year KS4. The aim of this was to allow a difference-in-differences analysis which will be discussed in a later section.

In order to inform selection to the impact analysis, the proforma also asked schools that have a three-year KS4 (for any or all subjects): which academic year they began the KS4 programme, and which term their Year 9 pupils started studying the KS4 curriculum: autumn, spring, or summer. Only schools that had a full additional year of the KS4 maths programme formed the three-year KS4 group (pupils that start KS4 in the autumn term of Year 9 and between the academic years 2007/2008 and 2018/2019). We excluded schools that had some pupils that started KS4 in Year 9 and some in Year 10 as we felt that this would dilute the 'treatment' effect because we could not obtain individual pupil data for just these pupils that did start KS4 in Year 9.

NFER sent a list of administrative data from schools (URN, school name, group allocation, and school-level data) to the DfE, which was used to match in KS2 and KS4 attainment data. NFER requested KS4 data for all pupils from participating schools between the years 2007/2008 and 2018/2019 and asked the NPD team to match in their KS2 attainment data for all 104 schools that were suitable for analysis: 32 (31%) schools delivered a three-year KS4 programme and 72 (69%) delivered a two-year programme.

In terms of pupil selection, we only selected pupils who had been at school for the full two or three years of KS4. Pupils who joined after the start of KS4 (September of Year 9 for three-year KS4 or September of Year 10 for two-year) may dilute the treatment effect as they could potentially move from a two-year KS4 school to a three-year KS4 school in Year 10. We identified whether a pupil has completed a full two or three years of KS4 by using the pupil date of entry variable, which is available in the NPD. We excluded 60,918 (~3%) pupil qualifications from analysis due to the pupil joining a school after the start of KS4.

Eligible cohorts for inclusion in the analysis

Although we exceeded the number of expected responses to the proforma, the pattern of response was quite different to what we anticipated. In particular, it was much more spread out in terms of year converted to a three-year KS4 (also see Figure 26 in the IPE section for more information). Our selection of schools was partly driven by the need to examine pre-programme trends, which meant that more than one pre-treatment cohort was needed. Originally we intended to limit the cohorts eligible for the impact analysis to three, however, the wider pattern of responses actually received meant that we needed to include more cohorts to make the analysis viable. Table 4 shows the five cohorts deemed eligible for the impact analysis (determined by the amount of pre- and post-event data available). This wider range of eligible cohorts brings added complexity to the analysis because of the changes to the GCSE system during that period (the method for conversion to a single GCSE scale is addressed in the Outcome Measures section below).

[5] There were two additional responses recorded: one school completed the survey twice so the duplicate response was removed and one independent school completed the survey: the survey was only open to state-funded secondary schools so its response was removed.

Table 4: Eligibility for event analysis (date of change to a three-year KS4)

Decision made (academic year)	Year 9s start studying KS4	First year of exams for 3-year KS4 (*c*+1)	Number of matched schools (three-year schools; two-year schools [never switchers]) (group number in the matched sample)	
2007/2008	2008/2009	2010/2011		*[not enough pre-event data available]*
2008/2009	2009/2010	2011/2012		
2009/2010	2010/2011	2012/2013		
2010/2011	2011/2012	2013/2014	27 (7; 20) (group 1)	eligible for event impact analysis
2011/2012	2012/2013	2014/2015	24 (7; 17) (group 2)	
2012/2013	2013/2014	2015/2016	22 (7; 15) (group 3)	
2013/2014	2014/2015	2016/2017	20 (6; 14) (group 4)	
2014/2015	2015/2016	2017/2018	11 (5; 6) (group 5)	
2015/2016	2016/2017	2018/2019		*[not enough post-event data available]*
2016/2017	2017/2018	2019/2020		data not available yet
2017/2018	2018/2019	2020/2021		
2018/2019	2019/2020	2021/2022		
2019/2020	2020/2021	2022/2023		

We realise that any decision to move from a two to a three-year KS4 will take a different amount of time in different schools. However for the purposes of this analysis, we are assuming that the final decision and move to implementation would take place in the academic year before the first Year 9 cohort start studying KS4.

One example shown in Table 5 is that a school informed us that it made the change in the academic year 2013/2014. The first relevant Year 9 cohort for this school, taking a three-year KS4 maths programme, would have started their KS4 studies in 2014/2015 and would have sat their maths GCSEs in the summer of 2017. A school that made the change a year later would see its first GCSE cohort in the summer of 2018. A school that made the change more recently, in 2017/2018, for example, was excluded from the impact analysis as its KS4 data would not yet be available. Furthermore, schools that, for example, only opened in the academic year 2016/2017 and, therefore, have no data prior to their opening were excluded from the impact analysis.

Table 5: Available cohort data for three-year KS4 schools

Group no.	2010/ 2011	2011/ 2012	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017	2017/2018	2018/2019
1	School decides to change	First cohort starting 3-year GCSE		First cohort take 3-year GCSE	Second cohort take GCSE	Third cohort take GCSE	Fourth cohort take GCSE	Fifth cohort take GCSE	Sixth cohort take GCSE
2		School decides to change	First cohort starting 3-year GCSE		First cohort take 3-year GCSE	Second cohort take GCSE	Third cohort take GCSE	Fourth cohort take GCSE	Fifth cohort take GCSE
3			School decides to change	First cohort starting 3-year GCSE		First cohort take 3-year GCSE	Second cohort take GCSE	Third cohort take GCSE	Fourth cohort take GCSE
4				School decides to change	First cohort starting 3-year GCSE		First cohort take 3-year GCSE	Second cohort take GCSE	Third cohort take GCSE
5					School decides to change	First cohort starting 3-year GCSE		First cohort take 3-year GCSE	Second cohort take GCSE

Matching process

Schools from the three-year KS4 arm were matched to schools from the two-year KS4 arm that were similar (in terms of the observable baseline characteristics listed below) at the time the decision was assumed to be made (that is, the year prior to the first Year 9 cohort starting KS4 study; Table 4). Matching was carried out on schools that operated a three-year KS4 for maths as this was the primary outcome. We note that it is a limitation of the secondary analyses that the match for all of the impact analysis is based on the length of the maths curriculum. (The composition of the final matched group is described in the Results section.) As noted above, we included multiple years - schools that changed to a three-year KS4 during the period 2011/2012 to 2015/2016 - and consequently we included a flag in the dataset to indicate which academic year the change to a three-year KS4 took place.

We used freely available school-level data to identify relevant characteristics that we felt it was reasonable to assume might be associated with the outcome of interest. During matching we used the following observable characteristics (as outlined in the study plan) of the schools that had introduced the change to a three-year programme until we obtained a matched dataset:

- school size (total number of pupils in school)
- school FSM (percentage of pupils eligible for FSM (school level))
- region;
- school type (type of secondary schools (middle, comprehensive to 16, comprehensive to 18, grammar, other) and;
- performance data (KS4 Attainment 8 performance)

The reason for this matching process was to carry out the difference-in-differences analyses using a list of two-year KS4 schools that were as comparable as possible to the three-year schools.[6]

Unlike RCTs where randomisation enables unbiased estimation of treatment effects for each observable characteristic, for quasi-experimental studies such as this, the assignment of length of KS4 is not random and the two groups being compared may be quite different in their observable characteristics as well as in their unobserved characteristics. There is a possibility of bias because the observed difference in GCSE maths scores between the two KS4 groups may also depend on characteristics that affect a pupil's GCSE maths score instead of being a result of the effect of the length of KS4 per se. Therefore, we felt it was necessary to include these variables in the matching process.

In addition to the variables identified above, we considered including the results of the proforma as variables within the matching process, in particular the factors affecting a school's decision to either change or keep the length of the KS4 programme. However, we found that these factors were not relevant to use in the matching process because the reasons for using a particular length of KS4 tended to clearly diverge by model (see IPE section for more information).
According to Little (2014), there are three main decisions affecting a matched dataset: the choice of measuring distance, the choice of matching strategy, and the choice of algorithm to perform matching. We opted to use propensity score as our choice of measuring distance. A propensity score is the probability of participating in a given intervention given a set of observable baseline characteristics. In our case, the relevant propensity is for a school to undertake a three-year programme of GCSE study in maths. We chose to estimate the propensity scores using a logistic regression model. The outcome of interest in the estimation of propensity scores is the binary indicator of whether a school is part of the main group of interest, that is, has a three-year programme of GCSE study. Our chosen matching strategy was a 1:5 matching strategy[7] without replacement using the 'nearest neighbour' matching algorithm. This meant that one school in the three-year KS4 arm was matched to, at the most, five schools in the two-year KS4 arm.

Matches were found using a combination of nearest neighbour and exact matching (by KS4 attainment and region) without replacement[8] using the MatchIt (Ho et al., 2013) package in R (R Core Team, 2017). The reason for choosing to match exactly by region and attainment was that these variables produced enough suitable matches that ensured 80% power. Initially we began with the matching strategy mentioned in the study plan, which was a 1:1 matching strategy using nearest neighbour matching and a caliper of 0.2. This resulted in only a small number of successful matches and we felt that the study would be significantly underpowered if we went with what was proposed in the study plan. We considered other matching strategies like optimal matching, coarsened exact matching, and full matching as well as a combination of nearest and exact matching to ensure we had enough three-year KS4 schools matched. By comparing the distribution of baseline covariates between the two groups of schools after matching obtained from different models, through data visualisations, we determined the most efficient matching model. The most successful matching model was the one that had the distribution of baseline covariates of the two-year group as similar as possible to the distribution of baseline covariates of the three-year group (Austin, 2011). A 1:5 match was selected as this resulted in a high number of schools that could be matched without reducing the matching quality. We also imposed no caliper as this resulted in more matches. Results of the matching process are found in the technical appendix (Appendix E).

In total we ran five matches, one for each year that a school with a three-year KS4 programme made a decision to change (group 1 to group 5 in Table 4). We began the matching process by selecting schools in the three-year KS4 group that made the decision in 2010/2011 and matched these schools to a list of all two-year KS4 schools (never-switchers). Once matching was complete, we then selected three-year schools in the three-year KS4 group that made the decision in 2011/2012 and matched these schools to a list of all two-year KS4 schools (never-switchers) excluding the ones the matched to three-years schools in the 2010/2011 group. We continued with this process until we matched schools in the three-year KS4 group that made the decision in 2014/2015 and matched this to a list of two-year KS4 schools (never-switchers) that did not match to any three-year KS4 schools in any group. It should be noted that this procedure effectively prioritised a close match for schools with an earlier three-year KS4 switch date. The procedure

[6] Analysis was carried out on three-year and two-year KS4 schools that responded to the proforma. It should be noted that these are not representative of schools across England.
[7] This differs from the study plan which mentions a 1:1 matching strategy. Using a 1:1 strategy we would not end up with enough schools to ensure the study was suitably powered. As a result of this, and as we had 159 unmatched comparison schools, we increased the number of two-year schools matching to a single three-year KS4 school to five. This means that schools in the three-year KS4 group matched to, at most, five schools in the never-switchers (two-year KS4) group.
[8] 'Without replacement' means that a school with a two-year programme can only match to one school with a three-year programme.

was also 'greedy' in the sense that multiple control schools could be matched to a single intervention school early on, making them ineligible for later matches. We were left with some unmatched three-year schools and in order to include as many of the schools in the analysis as possible, an additional match was carried out for these schools using 2016/2017 values of the matching variables (the next available academic year that had not yet been used in any matches). The reason for running this additional match was to ensure that we obtained 80% power. We conducted matches based on a treatment propensity score estimated from the following generalised linear model with a logit link function:[9]

$$logit(Group_i) = \beta_0 + \beta_1 Region_{i2015/16} + \beta_2 Schsize_{i2015/16} + \beta_3 KS4att_{i2015/16} + \beta_4 Schtype_{i2015/16} + \beta_5 FSM_{i2015/16} + \varepsilon_i$$

where (in the example of matching to a school that converted in 2015/2016):

- $Group_i$ is our 0/1 indicator of the length of KS4 programme for school i;
- $Region_{i2015/16}$ is a categorical variable indication the government office region in which school i is located in 2015/2016;
- $Schsize_{i2015/16}$ is a categorical variable indicating the quintile group school i falls into in terms of school size in 2015/2016;
- $KS4att_{i2015/16}$ is a categorical variable indicating the quintile group school i falls into in terms of KS4 Attainment 8 performance in 2015/2016;
- $Schtype_{i2015/16}$ is a categorical variable indicating the school type for school i in 2015/2016; and
- $FSM_{i2015/16}$ is a categorical variable indicating the quintile group school i falls into in terms of school-level FSM in 2015/2016.

When we restricted our sample of two-year KS4 schools using the matching process above, the balance between this sample and our sample of three-year KS4 schools improved for the variables that we chose to have an exact match (Region and KS4att). However, having disregarded those two-year KS4 schools that did not meet the exact matching requirement, the number eligible to be matched to each three-year KS4 school was generally five or fewer (that is, not greater than the maximum allowed by the 1:5 match). Nearest neighbour matching without a calliper (for Schsize, Schtype, and FSM) therefore had no effect on the match, in most cases, as selecting the five two-year KS4 schools that are the closest match to a particular three-year KS4 school requires at least six two-year KS4 schools to choose from. In effect, this meant that schools were matched by Region and KS4 attainment, but not the other three variables. The reason for choosing to match exactly by region and attainment was that these variables produced enough suitable matches to ensure 80% power. That said, there are likely, in hindsight, to have been alternative strategies that could have been employed, as discussed in the Study Limitations section.

The tables in Appendix E (see technical appendix) highlight the balance statistics prior to matching and after matching. Any remaining imbalance in Region and KS4 attainment was taken into account by including these as covariates in the regression models. As mentioned in Stuart (2010), this is the idea of 'double robustness' where regression adjustment is used to account for any residual imbalance between the groups. For the other three matching variables (School Size, School Type, and FSM) their inclusion as covariates in the regression is the only insurance against bias caused by an imbalance of these variables (not 'double robustness', as the matching was ineffective).

Outcome measures

The outcomes for this project were:

- primary outcome: GCSE maths performance;
- secondary outcome (1): GCSE English literature performance;
- secondary outcome (2): achieved five GCSE points between 5 and 8 including English and maths (see Table 6) from NPD; and

[9] As the model is the same for each year of change, we have provided the year 2015/16 as an example. If for a particular year data for variables were unavailable, data from the year prior to the change was used.

- secondary outcome (3): curriculum breadth.

Prior attainment measures

The prior attainment measure - a term more relevant in this context than 'baseline measure'[10] - for the primary outcome was KS2 maths attainment score. The variable KS2_MATTOTMRK was obtained from the NPD, which is the total marks achieved in maths tests (sum of Paper A, Paper B, and mental arithmetic tests).

The prior attainment measure for secondary outcome (1) was KS2 reading attainment score. The variable KS2_READMARK that was obtained from the NPD measured the total marks achieved in the reading test. For secondary outcome (2) of 5+ A*-C grades, we used a combination of KS2 reading and maths attainment scores as a baseline measure.

There is no baseline measure for the secondary outcome (3) of curriculum breadth.

Primary outcome

As outlined above, amendments to the design were necessary due to recruitment issues as a result of the Covid-19 pandemic. To remove any further burden on schools it was decided to use the data gathered and amend the design as there was sufficient data to maintain a similar MDES as the initial design for a single primary outcome of maths, but no longer enough for dual outcomes. The assumed ICC and correlation for this design would have required more responses for a single primary outcome of English or dual outcomes. Furthermore, due to changes in how GCSE performance is measured during the period of interest, we decided to use maths grades (rather than English) as the primary outcome because there is a higher level of consistency over the years in how this has been measured.11 Of the 252 schools that reported having a three-year KS4 for at least some subjects, 178 had a three-year KS4 for maths (also see Figure 3a).

The primary outcome was created using variables that computed the highest grade achieved in GCSE maths. The following NPD variables were used:

- KS4_GRADE; and
- KS4_MAPPINGDESCRIPTION.

Due to changes in the scoring system of GCSE grades between the years included in the analysis, the DfE published a table where GCSE grades achieved prior to the changes and after can be converted to the same scale (DfE, 2020, p. 40). We used this table to produce a single scale of GCSE maths points for the primary outcome (Table 6).

[10] Given that this analysis uses a difference-in-differences approach, we felt 'prior attainment' is a better description than 'baseline measures' as this refers to each school's pre-treatment (conditional) outcome data.

[11] Thought had been given to using Attainment 8 as the outcome measure but given changes to what is included within this measure it may be correlated with group membership and so it was decided not to use this as an outcome. Furthermore, recent changes in GCSEs (including, but not limited to, changes to the scoring system and changes to the equivalencies of some qualifications) meant Attainment 8 would not provide a single consistent measure over the time period this study required. There is also a higher correlation between prior attainment in maths (KS2 maths scores) and GCSE maths scores (correlation of 0.76) compared to prior attainment in English (KS2 English, reading) and GCSE English (correlation of 0.69) (EEF, 2013).

Table 6: GCSE points conversion table - used for 2017, 2018, and 2019 performance tables

Points awarded	1–9 GCSEs	Legacy GCSEs
8	9	A*
7.7	8	
7	7	A
6.3	6	
6		B
5.7	5	
5	4	C
4	3	D
3	2	E
2.5		
2		F
1.7		
1.5		
1	1	G

Source: DfE.

Secondary outcomes

Secondary outcome (1)—English literature GCSE

While a decision was made to use maths GCSE outcomes as the primary outcome to ensure a well-powered design and because of greater consistency in measurement, we also decided look at English literature GCSE outcomes as a secondary outcome (1). The reason for choosing English literature as opposed to English language was that there was a wider range of years of data available for English literature to carry out an analysis similar to the primary outcome analysis. Similar to maths, there have been changes to the scoring system for English GCSEs and we followed the same method as that adopted for maths in creating an outcome with a single scale (see Table 6). We used the following NPD variables:

- KS4_APENG—highest grade achieved in full GCSE, including double awards where appropriate, English (as used in the English threshold measures), 2005/2006 to 2012/2013;
- KS4_APENG_PTQ—highest grade achieved in full GCSE, including double awards where appropriate, English (as used in the English threshold measures), 2013/2014 only; and
- KS4_APENG_PTQ_EE—highest grade achieved in full GCSE, including double awards where appropriate, English (as used in the English threshold measures), 2014/2015 to 2016/2017.

Secondary outcome (2)— GCSE attainment across a broader range of subjects

To investigate the impact of the length of a KS4 programme on pupils' GCSE attainment across a broader range of subjects, a further secondary attainment outcome was analysed looking at the likelihood a pupil achieved at least five A* to C grades (or equivalent), including maths and English. As mentioned earlier, there is no single measure for this available from the NPD due to changes in the GCSE scoring system. As a result of this, we created our own measure by combining the following variables from the NPD:

- KS4_EXAMCAT—highest examination category achieved at GCSE and equivalent, 2004/2005 to 2012/2013;
- KS4_EXAMCAT_PTQ—highest examination category achieved at GCSE and equivalent, 2013/2014 only; and
- KS4_EXAMCAT_PTQ_EE—highest examination category achieved at GCSE and equivalent, 2014/2015.

We created a binary variable indicating whether a pupil achieved five GCSE points between 5 and 8 (see Table 6) including English and maths (and equivalent) or not. This is equivalent to whether a pupil achieved five A* to C GCSE

grades for pre-2017 data and whether a pupil achieved five 9 to 4 grades for data using the current GCSE scoring system.

Secondary outcome (3)—curriculum breadth

The third secondary outcome is curriculum breadth at KS4. Because this is a relatively novel outcome to explore in the NPD, it was necessary to explore the data (for example, whether qualifications could be meaningfully grouped, as in sub-outcome (c) below) before deciding which outcomes to use. This meant that details of the outcomes involved were unavailable in the statistical analysis plan and all the associated results must be considered exploratory to some degree. The aims in studying this outcome were (i) to identify whether a school's decision to move to a three-year programme of KS4 teaching is associated with a change in curriculum breadth and (ii) to identify whether there is a change in curriculum breadth over time for schools in general. Having considered previous literature (see for example Shapira and Priestley, 2018; Parameshwaran and Thomson, 2015; Gill, 2018) and the variables available in the NPD, it was decided that no single numeric outcome measure encapsulates what is understood by curriculum breadth. Instead, curriculum breadth is comprised of multiple related dimensions, including the number and diversity of qualifications taken by pupils, the number and diversity of qualifications made available within a school, and the uptake of subjects that do not contribute to the EBacc (for example, drama). In order to provide a more complete overview of these dimensions, the outcome of curriculum breadth was divided into four distinct sub-outcomes (a) to (d):

a) the number of KS4 qualifications entered per pupil;
b) the number of distinct KS4 qualifications entered per school in an academic year;
c) the overall proportion of all qualifications belonging to each subject area (for example, humanities) in an academic year; and
d) the overall proportion of pupils taking particular GCSE subjects (for example, French) in an academic year.

For both sub-outcomes (a) and (b), only distinct qualifications counted towards the total. This means that for (b) a qualification was counted only once per pupil cohort and the measure can be thought of as the number of available course choices for that cohort. Any differences in the subject or award label on the NPD led to two qualifications being treated as distinct. For example, 'full GCSE maths', 'full GCSE further maths', and 'BTEC maths' would count as three distinct qualifications, despite similarities between them. The quantities presented for sub-outcomes (a) and (b) were additionally displayed in terms of three categories:

- 'EBacc (core)' qualifications are qualifications in science (physics, chemistry, biology or combined), English (language, literature or combined), or maths that contribute to the English Baccalaureate.
- 'EBacc (optional)' qualifications also contribute to the English Baccalaureate but are not core, so largely consist of GCSEs in history, geography and languages but also includes some other subjects such as computer science.
- 'Non-EBacc' qualifications do not contribute to the English Baccalaureate.

Note that subjects such as religion and sports, while obligatory to teach at most schools, are not obligatory to take as qualifications and do not contribute to the EBacc; they are therefore 'Non-EBacc'.

Sub-outcome (c) was the overall proportion of all qualifications belonging to each 'subject area' in an academic year. These subject areas were broad categories of related KS4 subjects: 'English, maths, and science', 'humanities', 'languages', 'IT and computing', 'design and technology', 'arts', 'applied/other', and 'non-GCSE'. All qualifications were assigned to exactly one of these based on the subject and awarded labels given in the NPD. As this sub-outcome is concerned with subject areas, this does not directly align with EBacc eligibility. So, while EBacc-eligible qualifications were generally assigned to the 'English, maths and science', 'humanities', and 'languages' categories, this was not always the case. For example, EBacc-eligible computer science was classified as 'IT and computing' and EBacc-ineligible religion was classified as 'humanities'. Non-GCSE qualifications were categorised as 'non-GCSE', even when the subject itself matched another category (for example, 'BTEC maths' would be a 'non-GCSE', not 'English, maths and science'). GCSE qualifications that were of an applied nature (such as business studies) or did not belong in any of the other given categories (for example, sports) were categorised as 'applied/other'. Details regarding exactly which subjects were grouped into each subject area can be found in Appendix G in the separate technical notes document.

For sub-outcome (d), the overall proportion of pupils taking particular GCSE subjects in an academic year was calculated. In some cases differently-labelled GCSEs that were essentially the same subject were merged for this outcome, for example, GCSEs in 'expressive arts and performance studies' and 'drama and theatre studies' were both

treated as being the GCSE subject 'drama'. Details regarding which GCSEs were merged in this way can be found in Appendix G in the separate technical notes document. If at least one of the relevant labels was present in a pupil's records then they were counted as having taken that subject.

All sub-outcomes were based on qualifications entered, not qualifications obtained. Any qualification represented by a data row in the NPD was eligible for inclusion, regardless of the grade obtained and whether the course was completed. All qualifications were required to have been taken during KS4; any qualification with an exam date outside this period (two or three years depending on the school) was removed.

Although the impact analysis reported here is only able to look at curriculum breadth in KS4, there is some discussion of perceptions of the relative curriculum breadth and depth of KS3 and KS4 in the IPE section.

Sample size

Initially, the number of schools in the study was driven by sample size calculations, the primary outcome being a combination of GCSE English and maths scores. As indicated by the lack of prior studies in the introduction, there was limited prior evidence about the size of MDES, therefore the MDES was mainly informed by practical considerations. We proposed that a sample size of 100 schools within each school grouping would detect an estimated effect size of 0.15 with 80% power. This was based on the following assumptions: 180 pupils in a Year 11 cohort, an ICC of 0.2, and a correlation of 0.65 between combined English and maths KS2 and GCSE scores. This design was equally powered for analysis only on FSM pupils based on an assumption of 51 pupils per school who are flagged as ever having been eligible for FSM and using the same ICC and correlation.

As explained above, the survey period was disrupted by the Covid-19 pandemic and so we had to adjust our sample sizes for practical reasons. Originally, we were aiming for joint primary outcomes of maths and English. While updating the study plan, we reviewed the responses from the proforma and found that they indicated a different distribution of KS4 models to that initially predicted. Therefore, in discussion with the EEF, we decided that in order to obtain 80% power we would need to move to an unbalanced group design with a single primary outcome of GCSE maths scores. To detect an estimated effect size of 0.16, we would need 32 schools in the three-year KS4 programme (the 'treatment' group) and 72 schools in the two-year programme ('business as usual'). This was based on the following assumptions: 180 pupils in a Year 11 cohort, an ICC of 0.165 and a correlation of 0.76 between KS2 and GCSE. The correlation and ICC values were updated based on EEF guidance (EEF, 2013 and 2015).

The above estimates follow the standard EEF reporting framework for reporting power in randomised trials, which are not likely to be entirely accurate for the current difference-in-differences study. As an alternative we produced MDES estimates using the R package 'clusterPower' (Kleinman et al., 2021), which facilitates power calculations for simple difference-in-differences designs. This analysis was conducted post-hoc, was not specified in the study plan, and should be considered exploratory. Details are given with the main power calculations in the Results section.

Statistical analysis

Primary analysis

Change over time in the GCSE maths outcome pre- and post-treatment was displayed graphically to compare longitudinal trends between two- and three-year KS4 schools. This was done for both the unconditional or raw mean outcome and for the conditional mean. Inspecting the conditional pre-treatment trends in particular is useful for assessing the parallel trends assumption, which is central to the difference-in-differences design upon which the primary analysis is based. The parallel trends assumption is that the post-treatment trends would have been parallel in the absence of treatment, conditional on the covariates described for the primary analysis below. Placebo tests were also performed to assess the parallel trends assumption. This involves using only pre-treatment data and rerunning the primary analysis as if treatment began in a previous year in the data. If the resulting difference-in-differences estimate is statistically significant this suggests a violation of parallel trends, since there can be no actual effect of treatment in this case.

The primary outcome analysis was conducted at the pupil level, comparing GCSE grades in maths between schools with a three-year maths KS4 programme and schools with a two-year KS4 programmes. Pupil-level linear regression models with school-level random effects were run. Analysis looked to identify whether the long term levels of school performance had changed since the introduction of a three-year maths GCSE programme through the use of multilevel models. We are looking at averages over pre- and post-test periods here. We created a *year of conversion* variable to identify the year in which a cohort actually sat their GCSE examinations. For example, a school that started a three-year programme in the academic year 2014/2015 saw this first cohort sit their examinations in the summer of 2017. In this case, the 2016 GCSE cohort acted as the baseline year ($c = 0$) and years following this were recorded as *c*+1, *c+2*, and so forth. Pupils within comparison group schools received the same flag as pupils in their equivalent three-year KS4 school. The years preceding 2016 would be recorded as *c*-1, *c*-2, and so on. The positive years were combined into a single dichotomous variable to indicate the GCSE years since the change (*postchangeyears)*, as indicated in the regression equation below.

The regression model for the primary analysis is given by

$$Y_{ijt} = \beta_0 + \beta_1 \text{treatment}_j + \beta_2 \text{postchangeyears}_t + \beta_3 \text{treatment}_j * \text{postchangeyears}_t + \boldsymbol{X}'_{ijt}\boldsymbol{\beta} + U_j + \epsilon_{ij}$$

where

Y_{ijt} is the maths GCSE grade of a pupil i in group j at year t; and

treatment_j is a dummy variable set to 1 if a pupil is in the treatment group.[12]

Here postchangeyears_t is a binary variable taking the value 0 if t is prior to the first three-year KS4 cohort sitting their GCSEs and 1 otherwise. Although the comparison group consists solely of schools with a two-year KS4 for maths, and therefore did not make a change at any point, schools in the comparison group were assigned the same value for the postchange variable as their matched school. This was carried out in order to run a difference-in-differences[13] analysis and answer the research question: 'Do pupils attending secondary schools that teach KS4 over three years perform differently at GCSE to pupils in similar schools that teach KS4 over two years?'

The coefficients in the model were as follows:

- The average difference in GCSE maths attainment between the three-year KS4 group and the two-year KS4 group pre-conversion is identified by the coefficient β_1 (see 'Treatment' estimator in Table 13).
- The coefficient β_2 represents the average change in GCSE maths attainment post conversion compared to pre-conversion for the comparison group (see 'postchange' estimator in Table 13).
- The coefficient for β_3 identifies the average effect post policy change of a three-year programme at KS4 over what would have occurred to GCSE maths attainment scores had the treatment (converted to a three-year KS4) not been given (see 'difference in differences' estimator in Table 13).

Other covariates for this model ($\boldsymbol{X}'_{ijt}$) include: *gender* ('male' as the reference group), ethnicity ('white British' as the reference group), *SEN* ('identified SEN' as the reference group), and *pupil FSM* ('eligible for FSM' as the reference group) as binary variables; *academic year, school size*, *school FSM*, and *region*[14] as categorical variables and *KS2 prior attainment* as a continuous variable. We have included these covariates as it is likely that there will still be some variance

[12] Treatment group being schools undertaking a three-year programme of KS4 maths.
[13]There have recently been rapid developments in the difference-in-differences literature around the inclusion of time-varying covariates (Caetano, Callaway, Payne, & Rodrigues, 2022) and interpretation of estimates when treatment timing varies (Goodman-Bacon, 2021). Both of these areas of development are relevant to the current study. However, as this project was delayed for a considerable period due to Covid-19 and we were restricted by the data arrangements in place, we have been limited in the extent to which we have been able to retrospectively apply newer theories and techniques to the analysis.
[14] We did not have the school-level variables school size, school FSM, and region from the NPD. The available option was to use categorical variables, which we created for the matching.

left after selecting the comparison group of schools by matching eligible schools to the observable characteristics of the treatment schools.

U_j is a school-level random effect, which is assumed to be normally distributed, with mean zero a common variance amongst all schools. The residual error associated with the i^{th} pupil at year t is given by ϵ_{ijt} and this model was run in R using the lme4 (Bates et al., 2015) package with a full syntax trail.

Preferred method of analysis

We realised that there could be an argument for including school-level fixed effects in the model for this analysis and so ran checks to determine the approach. We ran a robust Hausman test to determine the preferred method of analysis for the primary outcome. If the null hypothesis is rejected, a fixed effect model still remains unbiased and consistent; by contrast, a random effect model will end up with biased and inconsistent estimates. The robust Hausman test showed a significant finding ($p < 0.05$) indicating that the fixed effect model is preferred. For the robust Hausman test we removed school-level fixed effects from the fixed effects model; we also did this for the fixed effects models described below. We then ran a sensitivity analysis comparing this fixed effects model to a random effects model with all variables present. Results of the sensitivity analysis showed very similar estimates for the variables of interest (postchange and difference-in-differences, see Appendix J). We also ran an additional sensitivity analysis comparing the results of the random effects model with all variables present to a correlated random effects model with all variables present. Results also showed very similar estimates for the variables of interest (also see Appendix J). Ultimately effect sizes were very similar and the random effects model was taken forward.

More detail is found in the section Choice of Statistical Method in Appendix J.

Further analyses

Exploration of heterogeneity over time

If results from the primary analysis indicated that the length of a KS4 programme had an impact on pupils' GCSE attainment in maths, then we planned to investigate in which year post-change the largest impact was detected. As our results did in fact indicate that such an impact existed, we investigated using the following steps.

The *year of conversion* variable was added to the model to determine any differential effects between the years after the change took place. For analysis purposes a number of dichotomous variables were created, one for each year. Interaction terms were introduced between these dichotomous variables and the group variable; the coefficients on these variables estimated the effect of interest. This follows a similar analytical design to Andrews et al. (2017) and their analysis of academies using the date of conversion. This design allowed us to see the trajectory prior to the introduction of the three-year programme and the trajectory after the change. With comparisons to the trajectory of schools maintaining a two-year programme, we were able to identify the differential effect of the new programme. The above analytical design can be considered a difference-in-differences design that explores heterogeneity over time.

Robustness of results to non-parallel trends

As will be seen in the Results section, it is doubtful as to whether the parallel trends assumption of a difference-in-differences design holds for the primary and secondary analysis of this study. One relatively simple extension of the primary analysis model is to add an interaction between treatment_j (see equation in the primary analysis section above) and a new covariate $\text{continuous\_years}_t$ - the number of years that have passed on the standardised timescale, starting at 0 in year c-9 (so c-8 becomes 1, c-7 becomes 2, and so forth). This model was calculated in order to allow for different linear trends between the two groups: since the parallel trends assumption seems untenable, this more flexible specification is more likely to produce a difference-in-differences estimate that reflects the actual causal impact of the three-year KS4 switch.

Another method for assessing the robustness of results to non-parallel post-treatment trends is described by Rambachan and Roth (2022) and is implemented using the R package 'HonestDiD' (Rambachan and Roth, 2021). This methodology begins with a model that assumes multiple pre-treatment years, one or more post-treatment years, and a single reference year (the final year before treatment: 'year 0', or in the language of this study, 'c'). The main parameters

of interest are interactions between each year and treatment group status. This is the same model specification as that described in the Exploration of Heterogeneity Over Time section above;15 the next steps which relax the parallel trends assumption use that model as a starting point.

In this methodology, post-treatment violations of parallel trends are allowed but are bounded by size constraints, which are specified as appropriate to the study. For this study, violations of parallel trends are restricted to the set:

$$\left\{ \delta : \forall\, t \geq 0 \; |\delta_{t+1} - \delta_t| \leq \bar{M} \times \max_{s<0} |\delta_{s+1} - \delta_s| \right\}$$

which is one of the potential constraints described by Rambachan and Roth (2022, p. 11). Here δ is the degree to which parallel trends are violated,16 s indexes the pre-treatment years, t indexes the post-treatment years, and $\bar{M}$ is a constant multiplier selected by the researcher. This constraint means, for $\bar{M} = 1$, that post-treatment violations of parallel trends (between two consecutive observation times) cannot exceed the maximum observed in the pre-treatment period. For $\bar{M} = 0.5$ they cannot exceed half the observed pre-treatment maximum violation, and so on, until $\bar{M} = 0$ permits no violation (so the original model, with parallel trends assumed). Confidence intervals for interaction terms from the original model can be compared with the wider intervals produced for a range of positive values for $\bar{M}$. For example, calculating the original model might produce a 95% confidence interval for the interaction between $year\ of\ conversion = c + 1$ and $group = 1$ (three-year KS4) of (0.05, 0.2). Using the methods above we might find that allowing for $\bar{M} = 0.4$ widens the interval to (0.01, 0.24), then allowing $\bar{M} = 0.5$ widens it further to (-0.01, 0.26). This provides a sort of 'tipping point': allowing for a maximum violation of parallel trends half that observed in the pre-treatment period ($\bar{M} = 0.5$) is the point at which the null value of 0 lies in the confidence interval.

As both pieces of analysis in this section were not specified in the analysis plan they must be considered exploratory.

Secondary analysis

To investigate whether the length of a KS4 programme had an effect on GCSE English literature scores[17] - secondary outcome (1) - we ran multilevel regression models using the lme4 package in R (Bates et al., 2015). This model was run using the same matched sample as identified in the primary analysis. We note that is it a limitation of the secondary analyses that the match for all of the impact analysis is based on the length of the maths curriculum. However, almost all (31 of the 32) 'three-year maths' schools also had a three-year KS4 for English literature, and all of the 'two-year maths' schools had a two-year KS4 for all subjects (including English literature).

The dependent variable for this model was GCSE English point score and covariates used in the model were the same as the covariates used in the primary model analysis apart from the prior attainment variable, which in this case was KS2 reading

To investigate whether the length of a KS4 programme had an effect on the examination category a pupil received - secondary outcome (2) - we ran a multilevel logistic regression model using the lme4 package (Bates et al., 2015) in R. This model was run using the same matched sample as identified in the primary analysis. The covariates in the model were the same as the covariates used in the primary model analysis apart from the prior attainment variable, which was a combination of KS2 English and maths scores. The dependent variable for this model was a binary variable indicating whether a pupil obtained five GCSEs with points between 5 and 8 (see Table 6) including English and maths (and equivalent) or not.

Curriculum breadth analysis

[15] It would of course be preferable to use the primary analysis model as a starting point as the robustness of these results is of principle interest. Unfortunately, this is not possible, to the best of our knowledge, as in that model the reference and pre-treatment years are not distinctly represented: the entirely period between c-9 and c is used as the reference level for the *postchangeyears* variable.

[16] δ represents the difference in trends that would have been observed in the absence of treatment. In the pre-treatment period δ is directly observed: it is the coefficient for the interaction between each pre-treatment year and treatment group status. A more precise description in causal terms is given by Rambachan and Roth (2022, p. 8).

[17] GCSE English literature scores, considered as GSCE English language scores between the years 2007/2008 and 2009/2010, were not available from the NPD.

Analysis performed on the third secondary outcome, curriculum breadth, consisted entirely of descriptive output displayed via graphs for all 104 schools in the primary impact analysis; no modelling framework was involved. This has changed since the study plan, which indicated that a measure related to curriculum breadth would be the outcome in a school-level linear model. There are several reasons for this change. First, there does not appear to be any precedent in previous literature for including curriculum breadth as an outcome in a statistical model. We considered the literature and felt that the multi-dimensional nature of curriculum breadth meant that it could not be easily summarised by a single number[18] and that at this stage, graphs were the best way to consider the different aspects involved in conjunction.

As detailed in the Outcome Measures section, the curriculum breadth outcome is comprised of four sub-outcomes - (a) the number of KS4 qualifications entered per pupil, (b) the number of distinct KS4 qualifications entered per school in an academic year, (c) the overall proportion of all qualifications belonging to each subject area in an academic year, and (d) the overall proportion of pupils taking particular GCSE subjects in an academic year. Graphs were produced displaying the overall trend amongst schools between 2007/2008 and 2018/2019 for each of these sub-outcomes. This range of academic years was decided by the analysis of the primary outcome as it allowed sufficient data before and after the switch to a three-year KS4 for schools that began this system between 2011/2012 and 2015/2016. Additionally, graphs comparing longitudinal trends for two- and three-year KS4 schools were produced for all sub-outcomes.

For graphs that compare two- and three-year KS4 schools, the timeline on the x axis ranges from '*c*-9 to *c*+6', where *c*=0 is the last year of a two-year KS4. This means the timeline has been standardised around the year of conversion and data for a particular year on this timeline has come from multiple academic years. '*c*+1' is the year when the first cohort under the three-year KS4 system take their Year 11 exams, that is, the first year we might expect to see an impact from the new system. Two-year KS4 schools have no literal 'year of conversion'; their value was assigned via propensity score matching to three-year KS4 schools with similar characteristics at the year of conversion, as described in the Methods section of the report. Because the year of conversion was not the same academic year for all schools (see Table 4), an increasingly small number of schools contribute data towards the far left (N = 5 three-year KS4 and N = 7 two-year KS4 schools at '*c*-9') and far right (N = 7 and N = 18 at '*c*+6') of the graphs. However, all schools contribute data in the range '*c*-5' to '*c*+2'; any apparent longitudinal trends outside this range may simply be due to a change in the composition of the groups themselves and should be treated with caution.

Each qualification was only counted for a single academic year, the year the pupil taking that qualification was in Year 11. Any qualifications taken by a pupil with an exam date before Year 11 were counted as occurring in Year 11, unless they were taken outside of the KS4 period, in which case they were discounted. Once a qualification was anchored to an academic year, it was consistently treated as having occurred in that year for all sub-outcomes (a) to (d). Note that this means that while sub-outcome (b) can be thought of as the 'number of qualifications offered' at a school, these are being counted when the cohort reaches Year 11, not when they were actually offered at the end of Year 8 or Year 9. For sub-outcomes (a), (b), and (c), short and double GCSEs were given a weight of 0.5 and 2, respectively, to better reflect their required time commitment and range of learning opportunities compared to a regular GCSE. For sub-outcomes (a) and (d) the mean number of qualifications per pupil (or proportion taking subjects for (d)) was calculated in one go, with all pupils pooled together; it was not first aggregated at school level and then aggregated again.

Missing data analysis

As per the study plan (Rutt et al., 2020), we envisioned that the number of pupils with missing outcome data would be small and these cases would be removed from the analysis without risk of bias. It was anticipated that the level of missing data would not exceed 5% at either the school or pupil level. As seen in the Impact Evaluation section of the report, there is a missing level of 0% at the school level. This was because de-identified pupil-level data from all schools was collected from the NPD after the matching stage. At the pupil level, there is a missing level of 0% for the primary outcome, which can be seen in Table 16 in the Impact Evaluation section of the report.

There is 8.5% missing data of the prior attainment variable, which is the only variable that passes the 5% missing threshold to carry out a missing data analysis. Generally, we would expect the reason for missing data to be unrelated

[18] It might be argued that each outcome could be included in its own statistical model, although this would still be problematic in some cases due to the categorical nature of the outcome (sub-outcome c) or the large number of separate outcome components involved (sub-outcome d).

to the group variable as the prior attainment variable is a statutory test (KS2 maths) and in most cases the KS2 tests would be taken in an entirely different school. A counter-argument would be that KS2 attainment could be missing due to the pupil having attended a private primary school: these institutions are not obliged to provide KS2 exam data. There might then (so this argument goes) be a link between the missingness mechanism and group status, if the socioeconomic context of the pupil's primary school carried through to their secondary school, which in turn influenced decisions about KS4 length. We think this argument worth mentioning, but as it relies on several steps of reasoning (some of which are fairly speculative) overall we favour the original statement: that group status and the missing data mechanism are unlikely to be related.

Since the pupil-level attrition for prior attainment was higher than 5%, it was important to explore the level of missing data and the extent of bias. To do so, we began by identifying the likely missingness mechanisms: missing completely at random (MCAR), missing not at random (MNAR), and missing at random (MAR). To test these assumptions we conducted diagnostics to establish any measurable predictors of missingness from the data. Initially, we looked for any imbalances between the groups (attrition and non-attrition) through cross-tabulations. If our groups are not equivalent (that is, statistically significantly different on any measure), carrying out a 'complete case analysis' (using only cases with complete data) may be biased as the study groups may not be representative of the original sample. This analysis also showed us whether cases with particular characteristics are more likely to have dropped out (biased attrition).

We ran logistic regression models with missingness on each variable of interest (1 = missing, 0 = otherwise) as the dependent variable and observable characteristics as independent variables. We ran five multilevel, multiple imputation models and compared results from a pooled multiple imputation model with the primary analysis. Variables used in the imputation models were the same as variables used in the primary outcome model.

Subgroup analyses

Subgroup analyses took place to explore the differential impact of the intervention when pupils' FSM status was taken into consideration. This was done using an interaction model that included the interaction of the difference-in-differences variable with pupil-level FSM. We also ran separate analyses of everFSM pupils as per the EEF requirement. We used the EVERFSM_6 variable from each spring school census between the years 2007/2008 and 2018/2019. These models were identical to those of the primary analyses except that they only included everFSM pupils.

We also explored the differential impact of the intervention when school deprivation was taken into consideration. This was done using an interaction model that included the interaction of the difference-in-differences variable with school-level FSM (used as a measure for school level deprivation). We used a quintile measure for school-level FSM for each school between the years 2007/2008 and 2018/2019.[19]

Additional analyses and robustness checks

To ensure that we obtained a sample of schools with a two-year KS4 programme that was as comparable as possible to schools with a three-year KS4 programme, we ran additional matching models with alternative matching methods. Alternative matching methods included changing the caliper setting and changing the matching strategy so that more than one school with a two-year programme matched to one school with a three-year programme. Results of this can be found in the technical appendix (Appendix E).

Additionally, we ran a range of models with each building on the previous version as additional sensitivity analyses to assess the robustness of findings of the primary analysis. The first contained treatment group variables, followed by pupil-level covariates, and a final model introducing school-level covariates. Results are presented in Appendix K.

Estimation of effect sizes

[19] A quintile measure for school-level FSM was used as this was the only measure of school-level deprivation available to us at the time of analysis.

As documented in the EEF's statistical analysis guidance (EEF, 2018), for comparability between EEF projects and with the wider literature, the EEF requires effect size calculations to be standardised. To do this, the following formula was used to calculate the effect size for a given regression coefficient estimate $\widehat{\boldsymbol{\beta}}$.

$$\mathbf{ES} = \frac{\widehat{\boldsymbol{\beta}}}{\sqrt{\boldsymbol{\sigma^2}}}$$

The numerator for all effect size calculations was the coefficient from the regression model adjusted for the baseline measure and other variables in the model. The denominator of the effect size was calculated using the square root of the sum of the within-cluster and between-cluster variance, which were extracted from a multilevel model without any covariates.

Confidence intervals for each effect size were estimated by using the equation:

$$\mathbf{CI}_{\boldsymbol{ES}} = \frac{\widehat{\boldsymbol{\beta}} \pm \mathbf{1.96} * \mathbf{SE}(\widehat{\boldsymbol{\beta}})}{\sqrt{\boldsymbol{\sigma^2}}}$$

where $\mathbf{SE}(\widehat{\boldsymbol{\beta}})$ is the standard error of $\widehat{\boldsymbol{\beta}}$.

Estimation of ICC

The intra-cluster correlation for the primary outcome was calculated using the following formula:

$$\frac{\sigma_{\mu}^2}{\sigma_{\mu}^2 + \sigma_{\in}^2}$$

where the numerator is the variability between schools and the denominator is the total variance, that is, the sum of the variance between and within schools. Both σ_{μ}^2 and $\sigma_{\in}^2$ were extracted from the multilevel model with no covariates.

An ICC score of zero implies no variation between schools and a high ICC score close to one implies high similarity between pupils' GCSE maths scores within the same school.

Implementation and process evaluation

The survey proforma described in the Impact Methods sections above also provided data for the IPE through a number of 'open' questions designed to gather reasons for particular models of KS4.

The qualitative evidence was collected via telephone interviews as it was not possible to conduct the case studies as originally intended owing to disruption caused by the Covid-19 pandemic. Instead of conducting up to 48 interviews across 12 case-study schools, it was agreed with the EEF to conduct up to 48 semi-structured telephone interviews with senior leaders.

The telephone interview sample comprised all the schools that had completed the online proforma and consented to be contacted for a follow-up interview. We contacted all of the schools that had given permission in four batches each containing 45 schools drawn to help ensure the interviews captured a range of KS4 model types and school characteristics, including school type and region. In order to be sensitive to the demands and pressures schools were facing, only a very light reminder strategy was implemented consisting of one short email.

The participating senior leader was selected in consultation with the school about who would be most appropriate to speak to. By the end of the interview period, 37 interviews with 40 individuals had been completed (three schools requested pair interviews involving the headteacher and a senior leader with responsibility for the curriculum). Table 19 in the implementation and process evaluation Results section provides a breakdown of the overall characteristic composition of the achieved interviews. The interviews took place in June and July 2021.

The interview schedule was developed using the research questions outlined above and was reviewed after four interviews. As part of this review, the order of some of the questions was changed and the wording of some questions clarified.

Analysis

The survey proforma was analysed using R. In contrast to the impact analysis, which was based only on the data for the 104 schools eligible for the primary outcome analysis, the survey frequencies reported in the IPE section are based on all of the schools that completed the proforma (n = 405). The 'open' questions were coded thematically by experienced researchers; coding frames were reviewed by at least one other researcher.

The qualitative data was analysed via MAXQDA using a top-level coding frame developed from the semi-structured interview schedule. After the first round of (deductive) coding, sub-codes were applied where needed to support the analysis of key themes using an inductive approach. As with the coding of the 'open' responses from the proforma, the coding frames and application of codes were quality assured by another experienced researcher. The data was also analysed by the length of KS4.

Table 7: IPE methods overview

Research methods	Data collection methods	Participants/ data sources	Data analysis methods	Research questions addressed
School proforma	Online survey	Sent to all state-funded secondary schools	Frequencies of closed questions (see Impact section), inductive coding of open questions	IPE_RQ3, IPE_RQ4
School interviews	Semi-structured interviews	37 telephone interviews	Mixed inductive/deductive coding; thematic analysis	IPE RQs 1-6

Timeline

The original timeline was delayed by four months due to Covid-19 after the launch of the survey, and by another three-four months due to the NPD application process and issues accessing the Secure Research Service (SRS).

Date	Activity	Staff responsible/ leading
November - December 2019	Draft study plan, advisory board meeting	NFER
w/c 20 January 2020	Project agreed	NFER
w/c 27 January	Project start-up, privacy notice Finalise proforma and confirm sample	NFER
February - April	Analysis of NPD data to determine performance matching variables	NFER
3 February	Agreed proforma entered onto online survey system, sample set up	NFER
w/c 24 February	Proforma (survey) sent to schools (after February half term)	NFER
March	Reminders to schools (email, social media)	NFER, EEF
	Survey paused due to Covid-19	
10 June	Survey relaunched (+1 reminder)	NFER, EEF
August - September	Initial descriptive analysis, PSM, match proforma data to NPD Review study plan for impact analysis in light of available data from proforma	NFER
September 2020	NFER and the EEF agreed to close the survey, with input from advisory group	NFER and EEF
October	Data cleaning	NFER
November	NPD request submitted	NFER
April 2021	NPD data made available to the project	NPD team
April - May	Run frequencies	NFER
April - August	Impact analysis	NFER
June	Contact interview sample (all those who expressed interest in doing a survey). No reminders to be sent. Split into two batches to mail out to manage the sample.	NFER
June - July	IPE—contact schools and conduct telephone interviews	NFER
July - August	IPE analysis	NFER
July - October	Report writing	NFER
December 2021	Draft report sent to EEF	NFER
Spring 2022	Advisory group review	NFER/EEF
Autumn 2022	Publication	NFER/EEF

Impact evaluation results

Participant flow including losses and exclusions

Once we received the information from the survey/proforma, we ran frequencies of the data to check whether the groupings in terms of length of KS4 we expected were present. The breakdown of schools by length of programme from the preliminary analysis is shown in Table 8. Just over a third of schools (35%) told us that they had a two-year KS4 at the time of the survey; that is, that pupils started studying the KS4 curriculum in Year 10. Pupils in the remaining two-thirds of schools started studying KS4 to some extent in Year 9. In most cases all pupils started studying KS4 in Year 9 (39%).

Table 8: Number of schools that responded to the proforma, by model of KS4

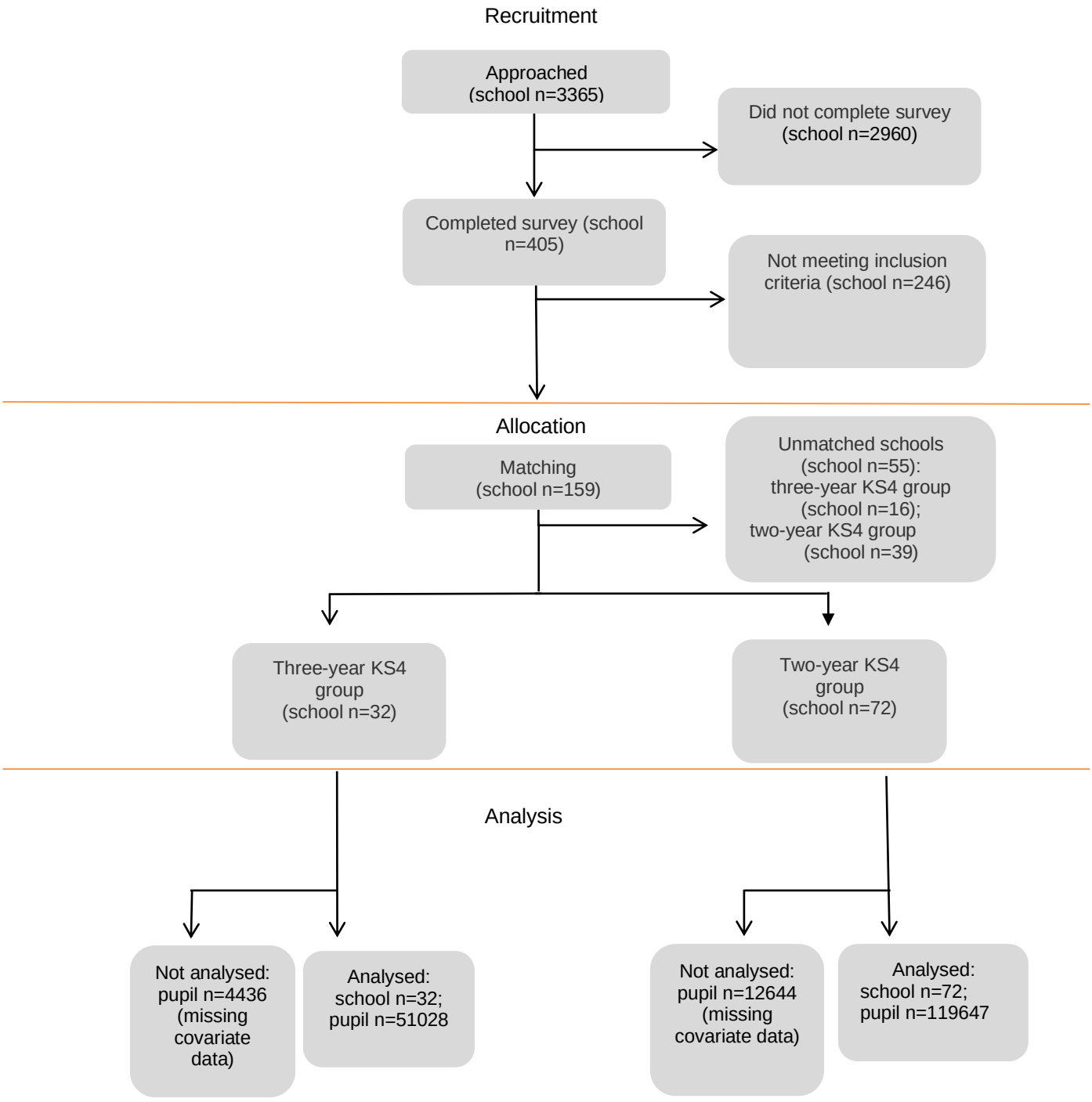
Length of programme	Frequency of schools (N)	%
All pupils start formal KS4 study of all subjects in Year 10	142	35
All pupils start formal KS4 study of all subjects in Year 9	157	39
All pupils start formal KS4 study of some subjects in Year 9 (and the rest of their subjects in Year 10)	95	23
Some pupils start formal KS4 study in Year 9 and some pupils start formal KS4 study in Year 10	11	3
Total	405	100

A single response item.
Responses include all schools that completed the proforma.

Figure 2 presents details of the participants' flow for the primary outcome through each stage of the study.

After excluding schools that made the decision to change prior to 2010/2011 and after 2014/2015 and schools that made the change in the spring or summer term, we had sufficient schools for a two-arm comparison: schools operating a three-year KS4 programme in at least maths, and schools with a two-year KS4.

Figure 2: Participant flow diagram for the primary outcome - GCSE maths performance



The flow charts below show how we obtained the number of schools eligible for matching. Figure 3a focuses on the flow of three-year KS4 schools from the survey stage to the matching stage and Figure 3b focuses on the flow of two-year KS4 schools from the survey stage to matching.

The starting number for Figure 3a includes schools in which (i) all pupils start formal KS4 study of all subjects in Year 9 (n = 157) and (ii) all pupils start formal KS4 study of some subjects in Year 9 (and the rest of their subjects in Year 10) (n = 95). This is because we selected schools into the three-year KS4 group according to the length of the maths KS4. Similarly, the 95 schools that started formal KS4 study of some subjects in Year 9 are also included in the starting number in Figure 3b, this time with all pupils starting formal KS4 study of all subjects in Year 10 (n = 142), because some of these schools have a two-year KS4 for maths.

Figure 3a: Selection of schools with a three-year KS4 for the matching stage

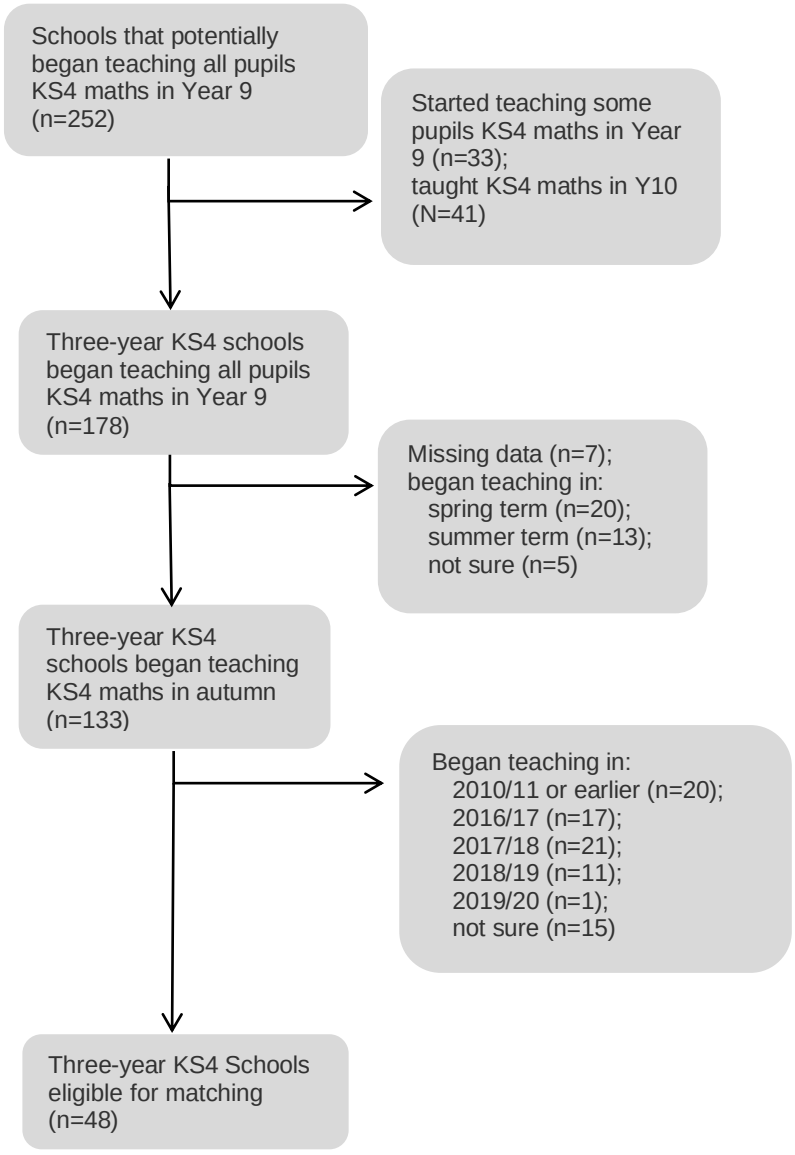


Figure 3b: Selection of schools with a two-year KS4 for the matching stage

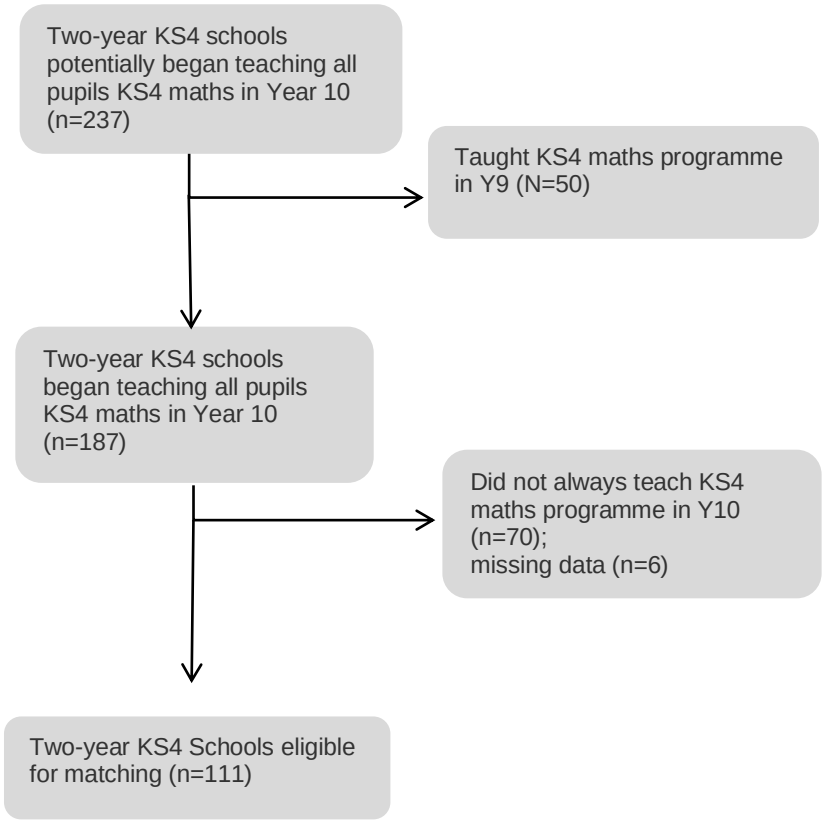


Table 9 provides details of the minimum detectable effect size at different stages in the study. The power formula used to calculate the MDES is given by Teerenstra et al. (2012). The pre-test/post-test correlation of the school-level cluster means ('r' in that paper) was approximated at the protocol stage using the pupil-level Pearson correlation between KS2 and KS4 maths scores. At the analysis stage 'r' could be calculated directly from the observed data, leading to higher values (0.91 versus 0.76) than originally predicted and consequently a lower MDES. Calculation of the ICC at the analysis stage is covered in the Methods section. The intra-cluster correlations (ICCs) and pre-test/post-test correlations at the protocol stage were obtained from EEF guidance (EEF, 2013 and 2015).

While the MDES at the analysis stage has been reported in Table 9 in accordance with the methods in the study plan, we also performed a further power analysis that we felt better reflected the current difference-in-differences study design using the `cpa.did.normal` function in the R package 'clusterPower' (Kleinman et al., 2021). An average cluster size of 1,641 was used: this is larger than the figure in Table 9 because it counts pupil GCSE maths results for all years in the study, not just a single year. Also unlike Table 9, it was specifically the post-switch difference-in-differences that was the target parameter for the MDES, not the difference in mean outcomes between the groups. Using an alpha value of 0.05, the MDES that could be detected with a power of 0.8 was found to be 0.10. An ICC of 0.18, an outcome variance of 2.68, and a cluster-level correlation between pre- and post-treatment maths scores of 0.92 were assumed for the calculation, all of which were derived from the observed data. For only FSM pupils the MDES was 0.14, using an average cluster size of 160, an outcome variance of 3.02, an ICC of 0.15, and cluster-level correlation between pre- and post-treatment maths scores of 0.85. This method has certain limitations: it assumed the 104 schools are equally divided into the control and intervention groups, as well as failing to include the explanatory power of baseline KS2 maths scores. However, despite these limitations we still present the above MDES figures as a (probably more accurate) alternative to those that populate the standard EEF template (Table 9).

Table 9: Minimum detectable effect size at different stages

		Study Plan		Analysis	
		Overall	FSM	Overall	FSM
MDES		0.16	0.17	0.11	0.12
Pre-test/post-test correlations	Level 1 (pupil)	0.76	0.76	0.73	0.73
	Level 2 (class)	-	-	-	-
	Level 3 (school)	0.76	0.76	0.91*	0.90
Intracluster correlations (ICCs)	Level 2 (class)	-	-	-	-
	Level 3 (school)	0.15	0.15	0.18	0.15
Alpha		0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8
One-sided or two-sided?		Two	Two	Two	Two
Average cluster size		180	51	136	14
Number of schools	Business as usual (2Y KS4)	72	72	72	72
	Comparison 1 (3Y KS4 all subjects)	32	32	32	32
	Total:	104	104	104	104
Number of pupils	Business as usual (2Y KS4)	12960	3672	9792	1008
	Comparison 1 (3Y KS4 all subjects)	5760	1632	4352	448
	Total:	18720	5304	14144	1456

* We note that the school-level pre-post correlation is higher than that observed in many educational studies. On further investigation, this figure is sensitive to the inclusion of schools with high mean KS4 maths scores; excluding schools with a mean score of six or more lowers the figure to be close to the pupil-level pre-post correlation. While 0.91 remains the best available estimate for the power calculation in this study, it may not reflect the figure amongst all English secondary schools.

Attrition

As this is not a trial but a QED using secondary datasets, we do not observe attrition in the traditional sense of schools or pupils dropping out of the evaluation. In terms of attrition due to absence of the primary outcome, no schools were lost as data was accessed, de-identified, from the NPD for their cohort of pupils. This was matched to a list of schools rather than pupils.

Attrition here may be considered in relation to pupils in analysed schools that ‘dropped out’ of the dataset. Pupils were lost from the analysis as they had missing covariate data (n = 17,080). Table 10 presents pupil-level attrition. On average,

9.1% of pupils were lost from analysis: 8.0% of pupils from three-year KS4 schools and 9.6% from two-year KS4 schools. The difference between the attrition rates in the two groups was reasonably small, especially given that treatment status was not a statistically significant predictor of KS2 maths scores being missing (see Missing Data section below). Where pupil data was missing it was baseline KS2 maths score in most cases; Table 16 provides a complete description of covariate missingness.

Table 10: Pupil-level attrition from the trial (primary outcome)

		Three-year KS4	Two-year KS4	Total
Number of pupils	Matching stage	55,464	132,291	187,755
	Analysed	51,028	119,647	170,675
Pupil attrition (from matching to analysis)	Number	4,436	12,644	17,080
	Percentage	8.00%	9.56%	9.10%

Source: NPD data 2007/2008 to 2018/2019.

Pupil and school characteristics

In total, 405 schools responded to the survey of which 104 were involved in the impact analysis. Table 11 presents key characteristics of the schools that formed the sample for the analysis of the primary outcome. Characteristics are given at baseline, or year 'c' (the year before the first three-year KS4 cohort take their GCSEs) where possible, or for 2018/2019 otherwise. As no schools dropped out, results of analysis to check balance at the matching stage is the same as at the analysis stage for school-level variables. As we did not obtain pupil-level data at the matching stage, we cannot say that results of analysis to check balance at the matching stage is the same as at the analysis stage for pupil-level variables.

Statistical tests (chi-squared or t-tests as appropriate) were used to check for any imbalance between two- and three-year KS4 schools, both at the initial response stage and the analysis stage; p-values from these tests are displayed in Table 11.[20] There were statistically significant differences in the N = 104 matched sample in school-level characteristics for urban versus rural status, Ofsted rating (in the baseline year), and percentage of FSM-eligible pupils (in the baseline year).[21] Of these, the most notable difference is in FSM eligibility as this variable was used in the matching process. As mentioned in the Methods section, we used propensity score matching to produce a list of two-year KS4 schools that were as comparable as possible to the list of three-year KS4 schools. This method attempts to reduce the bias due to confounding variables such as FSM eligibility, not eliminate confounding. To account for residual confounding, we included school-level FSM as a variable in all models analysed.

Additionally, we calculated the effect size for the difference (three-year minus two-year KS4) in continuous school-level characteristics. These are listed here in the format 'characteristic (standardised effect size, 95% CI)', average Attainment 8 score (-0.15, -0.58 to 0.27), percentage achieving A* to C including maths and English (-0.13, -0.54 to 0.29), percentage (school level) FSM eligibility (0.74, 0.33 to 1.16), and pupil count (0.15, -0.27 to 0.57).
There were statistically significant differences for all pupil-level characteristics (gender, white versus other ethnicity, FSM, and SEN status) indicating an association between these variables and length of KS4 programme. For the single, continuous, pupil-level characteristic of KS2 maths score, the standardised effect size was -0.06 (95% CI: -0.25, 0.13). As the effect size confidence interval straddles zero, this suggests no evidence of a difference in KS2 scores between the two matched groups.

[20] An equivalent table, but for all 405 schools that responded to the proforma survey can be found in Appendix I.
[21] Comparisons for all school-level characteristics were made using data from NFER's Register of Schools, outside of the SRS.

Table 11: Baseline characteristics of the groups as analysed (N = 104)

School level (categorical) [a]	National population 2018/2019 [b] [c]		Three-year KS4 maths group		Two-year KS4 maths group		Three vs two-year KS4
	n/N (missing)	%	n/N (missing)	%	n/N (missing)	%	P-value[d]
Urban or rural							
Urban	2902/3371 (6)	86%	>28/32 (0)	>88%	56/72 (0)	78%	0.047
Rural	463/3371 (6)	14%	X/32 (0)	X	16/72 (0)	22%	
School type							
Secondary school	3205/3371 (0)	95%	29/32 (0)	91%	69/72 (0)	96%	0.293
All through school	166/3371 (0)	5%	3/32 (0)	9%	3/72 (0)	4%	
Governance (2018/2019)							
Academy or free school	2600/3371 (35)	77%	25/32 (0)	78%	45/72 (0)	63%	0.117
Maintained	736/3371 (35)	22%	7/32 (0)	22%	27/72 (0)	38%	
Region							
North East	150/3371 (0)	4%	X/32 (0)	X	6/72 (0)	8%	0.991
North West/Merseyside	469/3371 (0)	14%	4/32 (0)	13%	11/72 (0)	15%	
Yorkshire and The Humber	330/3371 (0)	10%	X/32 (0)	X	X/72 (0)	X	
East Midlands	290/3371 (0)	9%	X/32 (0)	X	X/72 (0)	X	
West Midlands	390/3371 (0)	12%	3/32 (0)	9%	5/72 (0)	7%	
Eastern	380/3371 (0)	11%	4/32 (0)	13%	11/72 (0)	15%	
London	525/3371 (0)	16%	6/32 (0)	19%	9/72 (0)	13%	
South East	511/3371 (0)	15%	5/32 (0)	16%	13/72 (0)	18%	
South West	326/3371 (0)	10%	5/32 (0)	16%	12/72 (0)	17%	
Ofsted rating (baseline year)							
Outstanding			10/32 (0)	31%	18/72 (0)	25%	0.023
Good			13/32 (0)	41%	47/72 (0)	65%	
Requires improvement			9/32 (0)	28%	7/72 (0)	10%	
Inadequate			0/32 (0)	0%	0/72 (0)	0%	
School level (continuous)	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	P-value
Average Attainment 8 score (2018/2019)	2888/3371 (483)	43.4 (8.9)	31/32 (1)	49.5 (9.0)	70/72 (2)	50.8 (8.2)	0.492
Proportion of pupils achieving grades A*–C inc. English and maths (baseline year)			32/32 (0)	63.7 (13.3)	72/72 (0)	65.6 (15.4)	0.530
% FSM (baseline year)			32/32 (0)	16.6 (10.6)	72/72 (0)	10 (8)	0.001
Pupil count (baseline year)			32/32 (0)	1080.4 (349.6)	72/72 (0)	1030.5 (328.3)	0.485
Pupil level (categorical)			n/N (missing)	%	n/N (missing)	%	P-value
Gender (baseline year)							
Female			1915/3661 (0)	52%	3828/8261 (0)	46%	<0.001
Male			1746/3661 (0)	48%	4433/8261 (0)	54%	
Ethnicity (baseline year)							
White			2923/3661 (0)	80%	6771/8261 (0)	82%	0.007
Other ethnicity			738/3661 (0)	20%	1490/8261 (0)	18%	
FSM eligibility (baseline year)							
Eligible			568/3661 (0)	16%	772/8261 (0)	9%	<0.001
Not eligible			3093/3661 (0)	84%	7489/8261 (0)	91%	
Pupil has SEN (baseline year)							
Yes			673/3661 (0)	18%	1270/8261 (0)	15%	<0.001
			2988/3661 (0)	82%	6991/8261 (0)	85%	

No							
Pupil level (continuous)			n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	P-value
KS2 maths score (baseline year)			3661/3661 (0)	65.4 (21.2)	8261/8261 (0)	66.8 (20.6)	0.542

Source: NPD data 2007/2008 to 2018/19 and NFER's Record of Schools.
Counts lower than three had to be supressed for statistical disclosure control (indicated as X). If there was only one count that had to be supressed we also supressed the second lowest count so that the missing count could not be derived from the total.
'Baseline' is also referred to as '*c*' in the report.
Pupil-level characteristics are given for only those pupils that were included in the primary outcome analysis and who were taking their Year 11 exams in the school's baseline year.
(a) Time-dependent characteristics are given for either the 'baseline year' (the year before the first three-year KS4 exams) or for 2018/2019, as indicated in brackets in the left-hand column. All time-dependent characteristics are given for the baseline year, unless they were unavailable during the academic years corresponding to baseline.
(b) The national population is English secondary and all through schools, excluding independent schools. Population percentages may not add to 100 due to missing data.
(c) For variables where the characteristic is not fixed and can change over time, we have reported the status at the baseline time-point for each analysed school. As we have multiple cohorts in the analysis there is no single comparison year for the national population column and so these boxes are left blank.
(d) P-values were obtained by performing a chi-squared test (categorical data) or an independent samples t-test with equal variances assumed (continuous data), comparing characteristics in the two and three-year KS4 groups.

Statistical analysis results

Many of the plots in the forthcoming sections refer to the *year of conversion* timeline for this study, which is standardised relative to the treatment switch year. The *year of conversion* identifies the year in which a cohort actually sat their GCSE examinations. For example, a school that started a three-year programme in the academic year 2014/2015 saw this first cohort sit their examinations in the summer of 2017. In this case the 2016 GCSE cohort acted as the final pre-treatment year ('*c*') and years following this were recorded as *c*+1, *c+2*, and so forth. Pupils within comparison group schools were given the *year of conversion* of pupils taking GCSEs in the same academic year, at the matched three-year KS4 school. That is, while comparison schools had no literal year of conversion to a three-year KS4, the post-switch period was modelled as beginning in the same year as the intervention school they were matched to. The years preceding 2016 would be recorded in the format *c*-1, *c*-2 and so on.

Primary analysis

The primary analysis uses a difference-in-differences design, which makes the parallel trends assumption: that the mean trajectories of the two groups would have been parallel in the post-treatment period had the treatment (three-year KS4 switch) not occurred. As the truth of this assumption cannot be directly observed, it was inferred indirectly using the pre-treatment trends. We investigated this assumption using two methods: inspection of the mean pre policy switch GCSE maths score trajectories for the two groups and using placebo tests. These investigations were performed after the primary analysis model was calculated, but are presented before the primary analysis results due to their importance in interpreting model output.

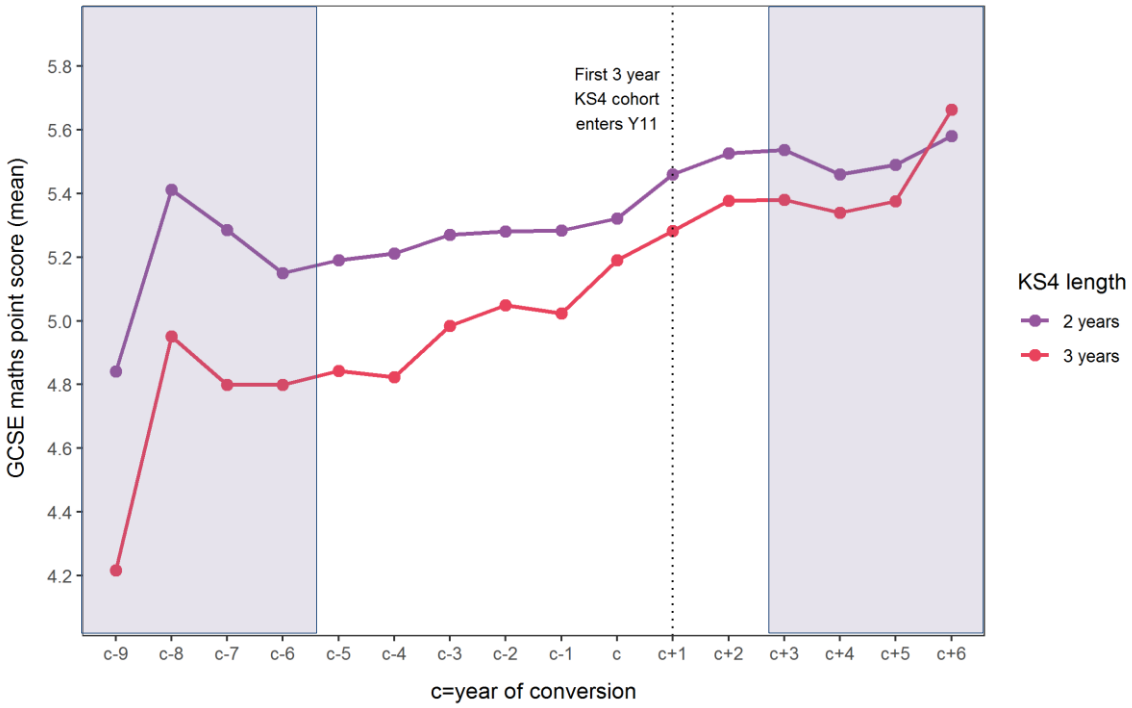
The following charts plot the mean GCSE maths point score, grouped by the length of KS4 programme and year of conversion for the unconditional (Figure 4a) and conditional (Figure 4b) specifications.[22] Both charts show that pupils in the two-year KS4 group perform better than the three-year KS4 group for all years in the study range. The key timepoints to consider are those between '*c*-5' and '*c*+2' because it is during this period that all analysed schools contribute data - this is indicated by the non-shaded area in the figures. Between the years '*c*-5' to '*c*+2' we can see that

[22] The conditional means are derived from the model used in the Exploration of Heterogeneity Over Time section below, with model covariates fixed at their mean (continuous covariates) or reference level (factors). Because the reference level of factors is being used, the absolute values of these means is not particularly meaningful, only the shape of the trajectories and whether the distance between the two groups varies over time. The same comments apply for the conditional plots for FSM pupils and for the English outcome below.

the gap between the two lengths of KS4 maths programme reduces over time, in both the conditional and unconditional plots. The most important assumption to ensure internal validity of the difference-in-differences model is the parallel trends assumption. However, we can see that the trends prior to conversion are not parallel, indicating that although the schools were similar in terms of Attainment 8 performance (through the matching process), there were some pre-existing differences in terms of maths performance trends over time. Specifically, these charts indicate that the three-year KS4 schools were on a faster upward trend than their matched two-year schools and therefore the trajectory that we see of both groups post conversion might be due to a continuation of the pre-policy performance (which was different for the two- and three-year schools), rather than the policy change itself.

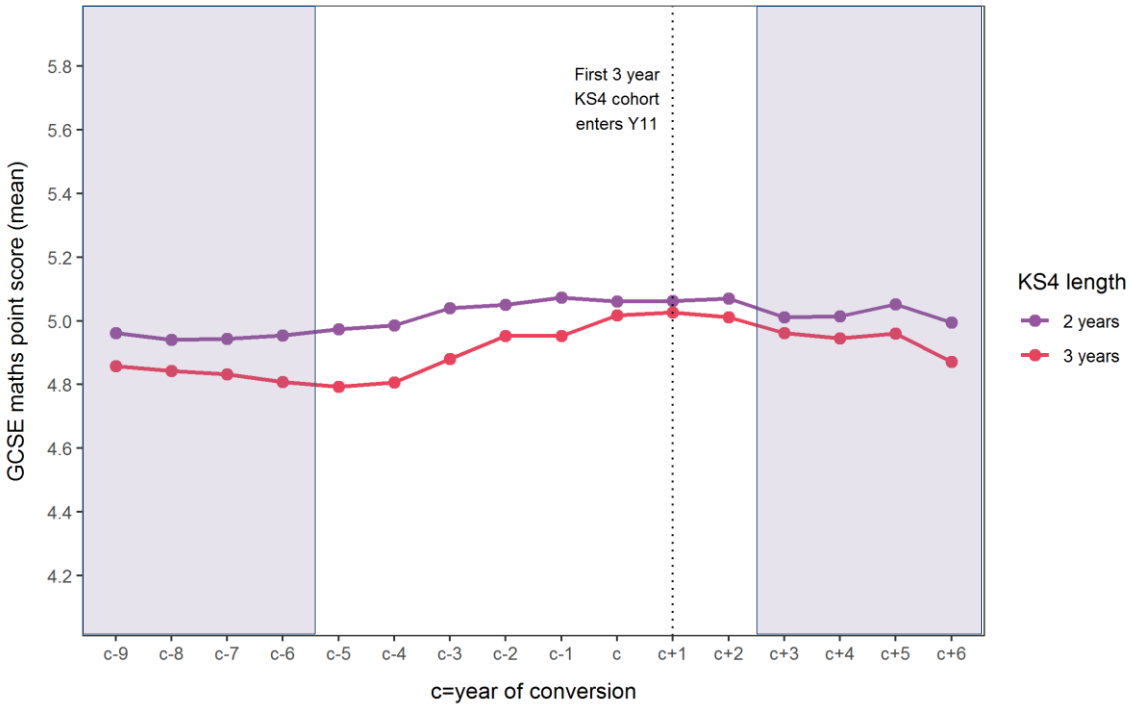
We note that at either ends of the plots, where the number of schools with data is much lower, we see some more extreme variation in the trends in the unconditional plot (Figure 4a). To illustrate why this is the case, Figure 4c shows the pattern for each group in terms of the mean GCSE maths point score over time, grouped by which academic year corresponded to the year of conversion '*c*' (the baseline year). As you can see, group number 4 that had the baseline year '2015/2016' was the highest performing group and so that led to a spike in the mean GCSE maths point score between the years '*c*-8' and '*c*-9'. GCSE maths scores generally increase over time for all groups in Figure 4c, though year-on-year there is substantial variation in the trend.

Figure 4a: Mean GCSE maths point score by length of KS4 programme and year of conversion—unconditional plot



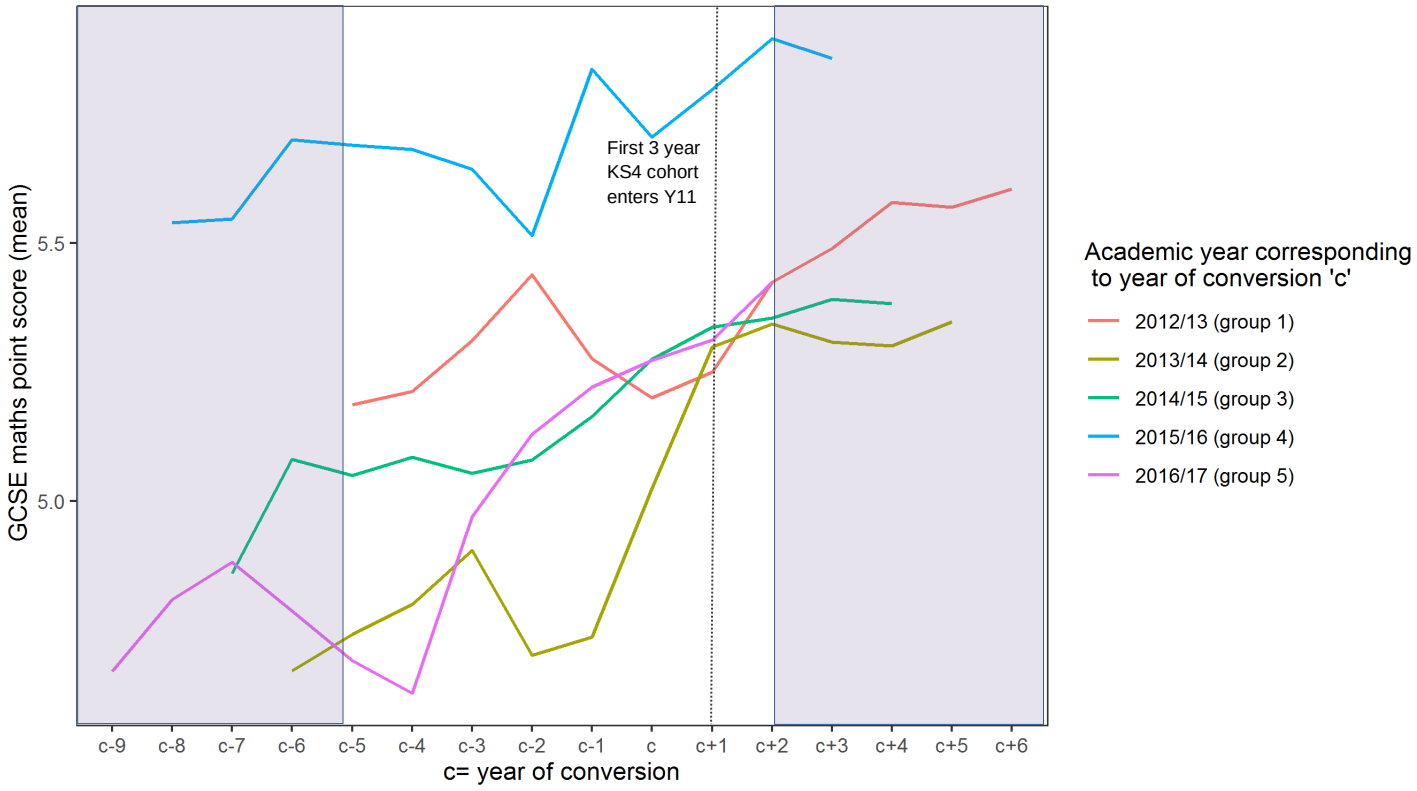
Source: NPD data 2007/2008 to 2018/2019.
Shaded areas indicate years with partial data.

Figure 4b: Mean GCSE maths point score by length of KS4 programme and year of conversion—conditional version



Source: NPD data 2007/2008 to 2018/2019.
Shaded areas indicate years with partial data.

Figure 4c: Mean GCSE maths point score by baseline year and year of conversion



Source: NPD data 2007/2008 to 2018/2019.
Shaded areas indicate years with partial data.

As described in the Primary Analysis part of the Methods section, placebo tests use the same model as the primary analysis (including covariates) but with two changes: only pre-treatment data (years *c*-9 to *c*) is used and the actual first year of treatment (*c*+1) is shifted back in time to a 'placebo' year in the pre-treatment period. In this case a series of eight models were calculated, corresponding to placebo years *c* to *c*-7 respectively (Table 12). If the difference-in-differences estimate is statistically significant for a placebo treatment year this indicates non-parallel trends in the pre-treatment period: there was no actual treatment that year, so the trajectories are converging (or diverging) for other reasons. As seen in Table 12, the statistical significance of the difference-in-differences estimate depends on the placebo year that is used. However, given that the estimates are significant for four placebo years, the general picture is that parallel trends do not hold in the pre-treatment period. The standardised effect sizes are small (absolute values between 0.001 and 0.048), but the difference-in-differences effect size for the primary analysis will also be seen to be small (0.046, see Table 13). That is, a violation of parallel trends in the post-treatment period of the magnitude observed in the pre-treatment period would potentially be sufficient to account for the entirety of the observed (apparent) effect of a switch to a three-year KS4.

Table 12: Difference-in-differences estimates for a series of models in which the actual first year of treatment is shifted backwards in time by one to eight years to a new 'placebo' year: c to c-7

Placebo treatment year *	N pupils		Difference-in-differences estimate	
	Pre-treatment (intervention, control)	Post-treatment (intervention, control)	Standardised effect size (95% CI)	p-value
c	82117 (24222, 57895)	11922 (3661, 8261)	0.015 (-0.01, 0.04)	0.250
c-1	69179 (20406, 48773)	24860 (7477, 17383)	0.023 (0.003, 0.042)	0.021
c-2	55449 (16342, 39107)	38590 (11541, 27049)	0.011 (-0.006, 0.029)	0.208
c-3	41302 (12255, 29047)	52737 (15628, 37109)	-0.004 (-0.022, 0.013)	0.624
c-4	26331 (7895, 18436)	67708 (19988, 47720)	-0.031 (-0.051, -0.011)	0.003
c-5	14334 (4370, 9964)	79705 (23513, 56192)	-0.045 (-0.07, -0.02)	<0.001
c-6	6516 (1880, 4636)	87523 (26003, 61520)	-0.048 (-0.084, -0.012)	0.009
c-7	1688 (461, 1227)	92351 (27422, 64929)	0.001 (-0.068, 0.07)	0.974

Source: NPD data 2007/2008 to 2018/2019.
* The year that is being modelled as if it was the first year of treatment (when the first three-year KS4 Year 11 cohort take their GCSEs). This actual first year of treatment, as used in the primary and secondary analysis, is '*c*+1'.

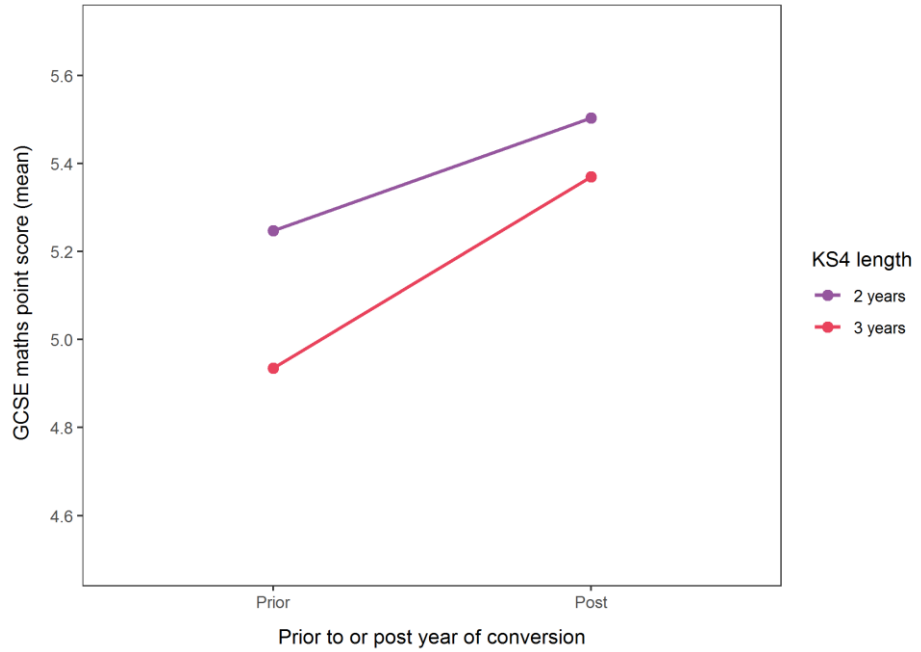
Having assessed the parallel trends assumption we now proceed to the main results for the primary analysis. These results should be interpreted in light of the investigations above: there is a strong possibility that some or all of the post-switch differences-in-differences in maths scores seen between the groups is due to a violation of the parallel trends assumption, not as a direct causal consequence of the three-year KS4 policy.

Figures 5a and 5b plot the mean GCSE maths point score, grouped by the length of KS4 programme and whether GCSEs were taken prior to or post the year of conversion. Figure 5a shows the plot of the raw or unconditional means while Figure 5b shows the conditional plot, including school- and pupil-level covariates. In terms of overall attainment, the raw plot in Figure 5a suggests that pupils in the three-year KS4 group improved more quickly than pupils in the two-year KS4 group.[23] The difference in raw mean GCSE maths point scores between 'prior to' and 'after' the year of conversion for the two-year KS4 group is 0.256 and for the three-year KS4 group is 0.435.. However, when we control

[23] As described in the Methods section, schools in the two-year KS4 group were matched to a list of three-year KS4 schools using KS4 Attainment 8 progress scores. This means that schools in these groups may have differed in GCSE mathematics point scores as schools were not matched using these variables.

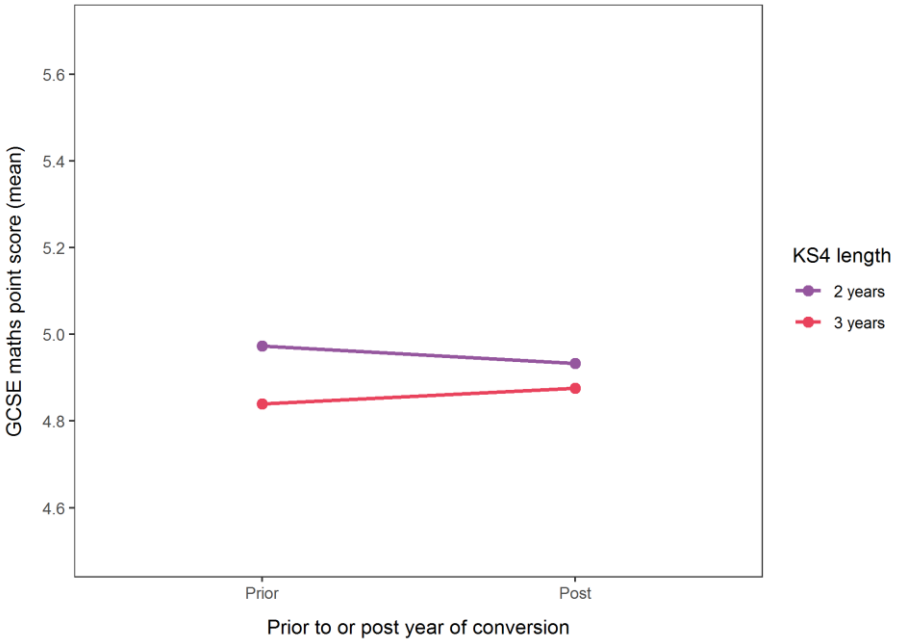
for school- and pupil-level characteristics (Figure 5b) a different picture emerges: while the three-year group's attainment still appears to be increasing, the two-year KS4 group's attainment is decreasing over time.

Figure 5a: Mean GCSE maths point score by length of KS4 programme and whether the year is prior to or post the year of conversion—unconditional plot



Source: NPD data 2007/2008 to 2018/2019.

Figure 5b: Mean GCSE maths point score by length of KS4 programme and whether the year is prior to or post the year of conversion—conditional on covariates



Source: NPD data 2007/2008 to 2018/2019.

We adopted a difference-in-differences research design to compare the difference in GCSE maths outcomes post policy change between schools that are in the three-year KS4 group and two-year KS4 group. The purpose of this design is to control for pre-treatment differences in levels between the groups and pre-treatment trends. To further account for pupil- and school-level characteristics, we implemented this design using a multilevel linear regression model. The choice of analysis method for choosing the most appropriate model is explained in Appendix K. Table 13 reports the results of the regression model for the primary outcome (maths), controlling for pupil- and school-level characteristics.

Table 13: Primary analysis results

		Length of KS4 programme		Effect size		
		Three-year KS4 group	Two-year KS4 group			
Outcome	Variable of interest	n (missing)	n (missing)	Total n (intervention; control)	Standardised effect size (95% CI)	p-value
GCSE maths scores	Treatment	51028 (0)	119647 (0)	170675 (51028; 119647)	-0.081 (-0.160, -0.003)	0.045
	Postchange	51028 (0)	119647 (0)	170675 (51028; 119647)	-0.025 (-0.039, -0.01)	0.001
	Difference-in-differences	51028 (0)	119647 (0)	170675 (51028; 119647)	0.046 (0.032, 0.060)	< 0.001

Source: NPD data 2007/2008 to 2018/2019.

To understand the context of our results, we compare pupils' GCSE maths scores between the two-year and three-year groups pre-conversion, which is reported by the coefficient on the 'treatment' variable. The estimate for this effect is moderate in size but negative and significant indicating that before converting to a three-year programme, schools in the three-year KS4 group performed (on average) lower in GCSE maths compared to schools in the two-year KS4 group during the same period.

Next, we consider the maths performance of pupils in schools that maintained a two-year KS4 throughout the time-period to provide an estimate of what would have happened to our treatment schools in the absence of the policy change to a longer KS4. To compare pupils' GCSE maths point scores pre and post conversion for two-year KS4 schools, we look at the variable *postchange*.[24] This is, again, very small but both negative and statistically significant indicating that for two-year KS4 pupils, GCSE maths point scores were *lower* during the post-policy-change period compared to the pre-policy change period, once we controlled for pupil and school characteristics; this can also be seen in the conditional plot in Figure 5b. This suggests that even in the absence of the policy change (longer KS4), there was a small negative change to average maths scores over time after controlling for pupil and school characteristics.

The next step was to investigate whether there is a differential change between the groups (two-year KS4 vs three-year KS4) between the pre- and post-policy-change periods. Here, we are particularly interested in the difference-in-differences variable; this interaction captures the difference in GCSE maths score change between three-year and two-year schools after the policy switch (to a longer KS4). The coefficient for the difference-in-differences interaction therefore seeks to capture the impact of the policy change on GCSE maths scores, estimating the average effect of treatment on the treated (ATT) (see Table 13). Results showed a very small significant difference and confirmed the finding suggested above: the cohort of pupils in the year after the three-year KS4 group implemented this policy saw a larger improvement in their GCSE maths compared to performance in their school in the previous year and what would have been expected in the absence of this policy change (as proxied by schools that did not make this change) when pupil and school characteristics are taken into account. However, as noted above, the difference of groups in terms of maths performance pre-policy change means that this finding needs to be interpreted with caution as it is quite possibly due to pre-policy change differences in the trajectory of the two groups.

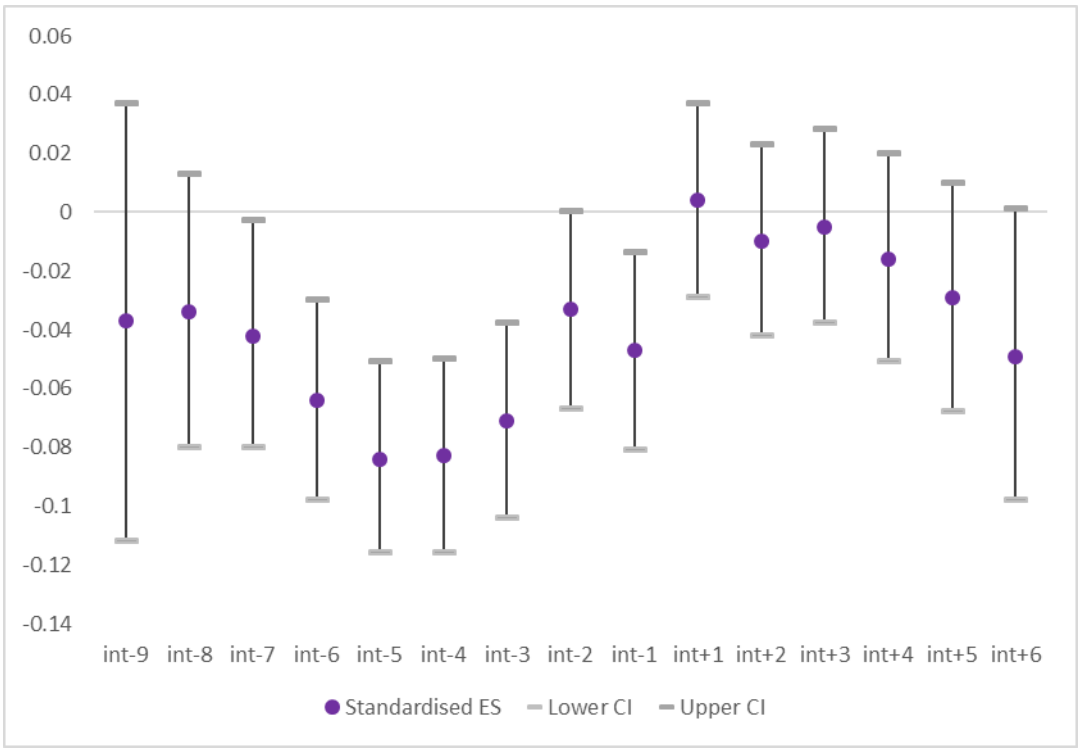
[24] *Postchange* is a binary variable taking the value zero if a pupils' GCSE mathematics point score was awarded prior to the year in which the matched schools converted to a three-year KS4 programme and one if a pupils' GCSE mathematics point score was awarded after the year in which matched schools converted to a three-year KS4 programme.

Exploration of heterogeneity over time - maths

We proceeded to investigate whether the (potential) impact of a three-year KS4 varied year-on-year in the post-treatment period using the framework described in the Methods section (Exploration of Heterogeneity Over Time).

Figure 6 displays the 'int' variable, which is the interaction between treatment group (two- or three-year KS4) and year (*c*-9 to *c*+6, where *c* is the reference level).[25] For instance, 'int+1' indicates how much GCSE maths scores have diverged between the two- and three-year KS4 groups in the first year post policy switch *compared to reference year 'c'*. This final piece of context is important and explains why 'int+1' to 'int+6' are not statistically significant (or even generally positive) estimates, in contrast to the positive and statistically significant difference-in-differences estimate from the primary analysis. Difference-in-differences estimates appear to be highly dependent on the choice of reference period: the final pre-treatment year here versus the entire pre-treatment period in the primary analysis. Note also that the difference between 'int' estimates in consecutive pre-treatment years (for example, int-5 to int-4) quantifies the violation of parallel trends between those years. The largest pre-treatment violation actually occurs between *c*-1 and *c* when the schools had already implemented the three-year KS4 system but the first cohort to have started KS4 in Year 9 had not yet taken their GCSEs. This may point to changes occurring in the Year 10 curriculum or in the school's administration generally that were related to the extended KS4 length.

Figure 6: Exploration of heterogeneity over time - maths, interaction of year (c-9 to c+6, where c is the reference level) with treatment



Source: NPD data 2007/2008 to 2018/2019.

Robustness of results to non-parallel trends

Given the probable violation of the parallel trends assumption for the primary analysis, another (potentially more credible) formulation of that model is to add two further covariates: 'Continuous Years' and an interaction between Continuous Years and Treatment (Table 14). Here Continuous Years is the number of years that have passed on the standardised (relative to the 'year of change') timescale, treated as a continuous variable (so year *c*-9 becomes 0, *c*-8 1, *c*-7 becomes 2, and so on). Although maths GCSE scores begin lower in the three-year KS4 group, the positive interaction between Continuous Years and Treatment allows these scores to converge on those of two-year KS4 schools by a fixed amount each year: a linear deviation from parallel trends. As in the primary analysis, the main parameter of

[25] Also see Appendix M for regression model coefficients.

interest is the difference-in-differences effect size estimate of 0.024 (95% CI: 0.000, 0.049), which is smaller than the estimate from the primary analysis and is now not statistically significant.

If the interaction between Continuous Years and Treatment in the post-treatment period is assumed to be a continuation of the pre-treatment convergence (that is, a violation of parallel trends and not because the treatment is actually becoming more effective each passing year) then this difference-in-differences term can be viewed as the 'effect' of the three-year KS4 switch.

Table 14: The primary analysis model expanded to allow a linear deviation from parallel trends

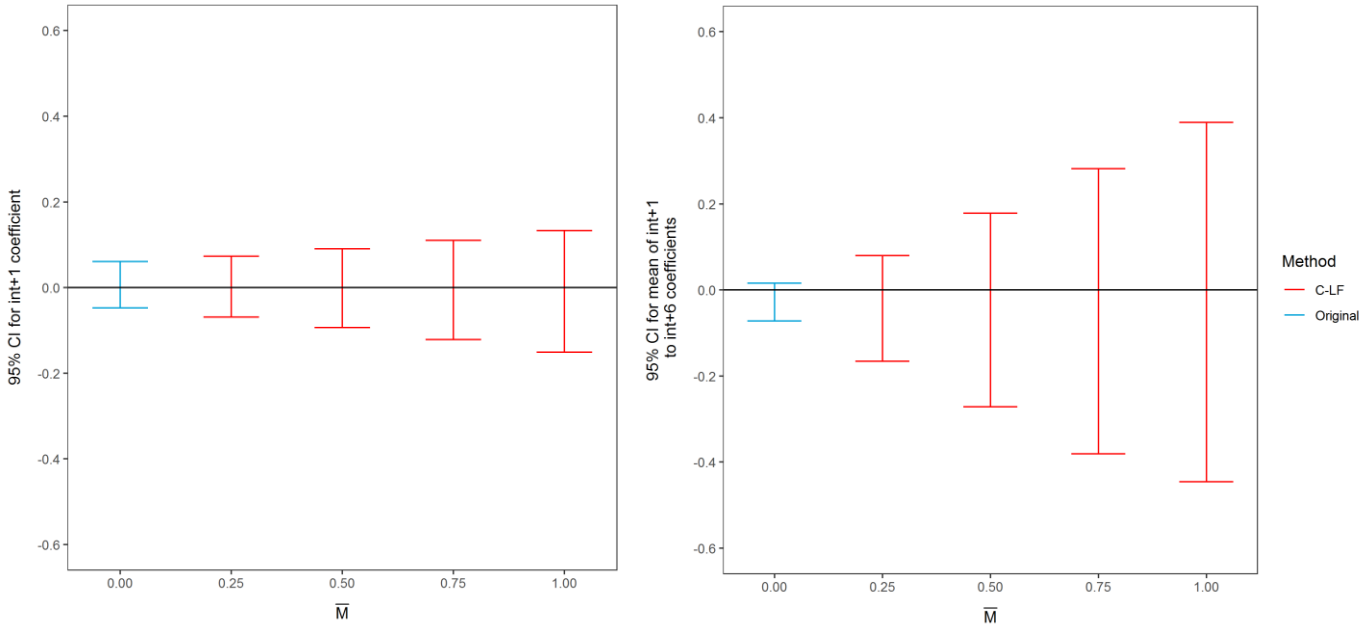
		Length of KS4 programme		Effect size		
		Three-year KS4 group	Two-year KS4 group			
Outcome	Variable of interest	n (missing)	n (missing)	Total n (intervention; control)	Standardised effect size (95% CI)	p-value
GCSE maths scores	Treatment	51028 (0)	119647 (0)	170675 (51028; 119647)	-0.1 (-0.181, -0.019)	0.017
	Postchange	51028 (0)	119647 (0)	170675 (51028; 119647)	-0.017 (-0.034, -0.001)	0.032
	Difference-in-differences (Treatment*Postchange)	51028 (0)	119647 (0)	170675 (51028; 119647)	0.024 (0.000, 0.049)	0.054
	Continuous Years	51028 (0)	119647 (0)	170675 (51028; 119647)	0.003 (-0.028, 0.034)	0.830
	Treatment*Continuous Years	51028 (0)	119647 (0)	170675 (51028; 119647)	0.004 (0.000, 0.007)	0.033

As detailed in the Methods section, we also investigated the robustness of the confidence intervals seen in Figure 6 above to violations of the parallel trends assumption.[26] Specifically, we allowed for post-treatment violations no larger than $\bar{M}$ times the largest observed violation in the pre-treatment period, for a range of values of $\bar{M}$ (Figure 7). Results are presented for the int+1 (treatment = 'three-year KS4' by year = *c*+1) interaction coefficient and also the mean of the post-treatment interaction coefficients int+1 to int+6 (the 'average effect' of a three-year KS4 in those six years). Estimation of these robust confidence intervals was performed using the 'conditional least-favourable hybrid' (C-LF) method recommended by the HonestDiD package authors (Rambachan and Roth, 2021).

The methods used in this robustness analysis are more useful when the confidence interval from the original model excludes zero: they then investigate at what point confidence intervals for increasing values of $\bar{M}$ do contain zero. However, we considered that it was still worth presenting this analysis in order to better estimate the degree of uncertainty around estimates in the absence of parallel trends. As might be expected, the original confidence interval for int+1 of (-0.048, 0.061) widens as $\bar{M}$ increases, to (-0.151, 0.133) at $\bar{M} = 1$. The confidence interval for the mean of int+1 to int+6 widens far more, from (-0.072, 0.015) originally to (-0.446, 0.390) at $\bar{M} = 1$, reflecting the additional uncertainty introduced by non-parallel trends in each consecutive year in the post-treatment period.

Figure 7: 95% confidence intervals for the int+1 coefficient (left plot) and for the mean of the post-treatment coefficients int+1 to int+6 (right plot)—confidence intervals are given for a range of values of $\bar{M}$, where post-treatment violations of parallel trends are constrained to $\bar{M}$ times the observed pre-treatment maximum

[26] Unlike elsewhere in the report, unstandardised regression coefficients are used here by necessity, which is why the 'original' interval for int+1 on the left-hand plot of Figure 7 is not on the same scale as the int+1 interval on Figure 6.



Source: NPD data 2007/2008 to 2018/2019.

Subgroup analyses

Two subgroup analyses (model 1 in Table 15a and model 3 in Table 15b) were conducted to explore the differential impact of the length of KS4 programme on pupils' GCSE maths point scores for pupils that have ever been eligible for FSM ('everFSM').

Another interaction model (model 2 in Table 15a) explored the differential impact of the length of KS4 programme on GCSE maths point scores when school-level FSM was taken into consideration. Results from the interaction models are summarised in Table 15a and Table 15b.

In model 1, pupil-level everFSM was interacted with the difference-in-differences term, whereas in model 2 school-level FSM was interacted with the difference-in-differences term.

We also ran separate analyses of everFSM pupils as per the EEF's requirement (model 3; Table 15b). This analysis is similar to the analysis for the primary outcome and as such we start that section with an exploration of the pre-treatment trends.

Eligibility for FSM - analysis including all pupils

Results from model 1 (Table 15a) suggest that when comparing everFSM eligible pupils to those not eligible for everFSM, switching to a three-year KS4 maths programme did not have a statistically significant differential effect on GCSE maths scores.

Table 15a: Subgroup analysis

Model number	Outcome	Variable of interest	Effect size (95% CI)	p-value
1	Pupil-level everFSM	Difference-in-differences * everFSM (pupil level)	0.034 (-0.009,0.076)	0.119
2	School-level FSM	Difference-in-differences * FSM (school level, lowest band)	0.186 (0.115,0.256)	<0.001
2	School-level FSM	Difference-in-differences * FSM (school level, 2nd lowest band)	0.114 (0.06,0.168)	<0.001
2	School-level FSM	Difference-in-differences * FSM (school level, middle band)	-0.008 (-0.06,0.045)	0.777
2	School-level FSM	Difference-in-differences * FSM (school level, 2nd highest band)	0.139 (0.082,0.197)	<0.001

Source: NPD data 2007/2008 to 2018/2019.

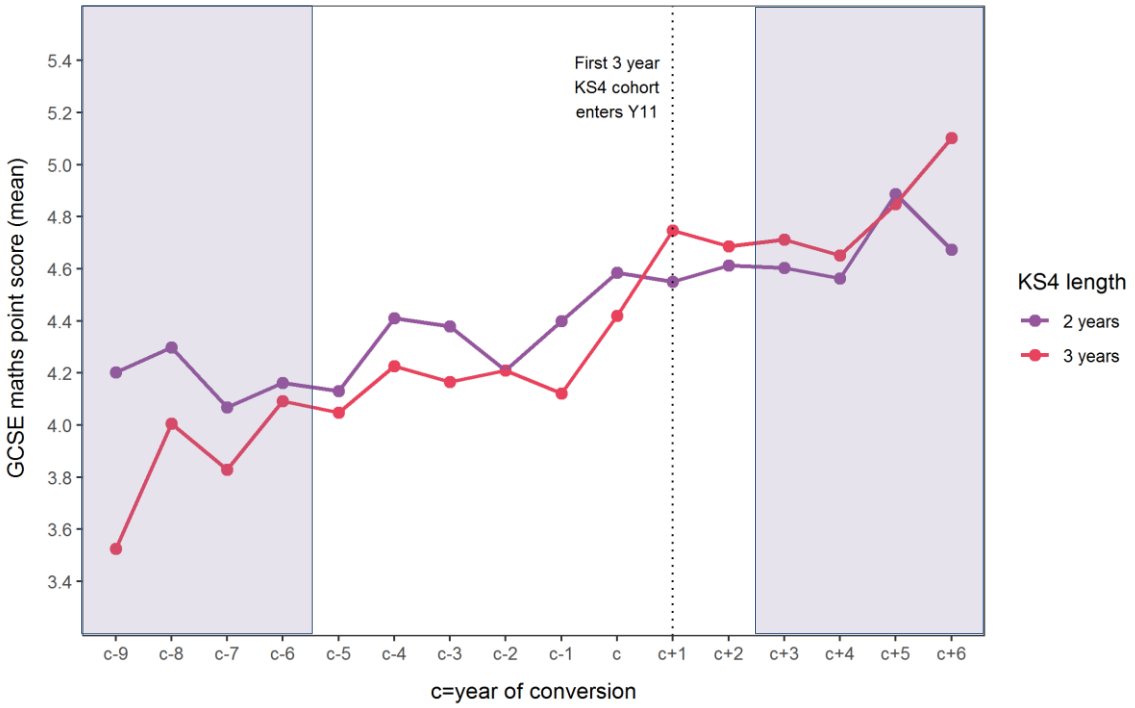
Model two (Table 15a) compared pupils in the highest FSM band (top 20% of schools analysed with the highest percentage of FSM) to schools in each of the other bands.[27] The results found statistically significant evidence to suggest that pupils in the three-year KS4 group in the lower FSM bands improved more quickly in their GCSE maths over time compared to pupils in the three-year KS4 group in the highest FSM band (relative to pupils in two-year KS4 schools) when pupil and school characteristics were taken into account. This is with the exception of schools in the middle band where there was no statistical difference.

Eligibility for FSM - 'everFSM' pupils only

Here we follow a similar structure to that of the primary analysis findings: first, we present charts showing the mean GCSE maths score for everFSM pupils only before presenting the findings of the model.
The following plots show the mean GCSE maths point score grouped by the length of KS4 programme and year of conversion for FSM eligible pupils for the unconditional (Figure 8a) and conditional (Figure 8b) specifications. The key timepoints to consider are those between '*c*-5' and '*c*+2' because it is during this period that all analysed schools contribute data - indicated by the non-shaded area in Figure 8a and Figure 8b. Between the years '*c*-5' to '*c*+2', we can see that the gap between the two lengths of KS4 maths programme fluctuates along a similar path in both the unconditional and conditional versions and visually, at least, it appears that the performance of the everFSM pupil subgroup is similar across the two-year and three-year schools. As such, the parallel trends assumption appears to hold in this specification. However, this is taken in the context of a subgroup analysis within the main analysis for which the parallel trends have been violated. We believe the caution in interpreting the primary analysis should therefore also be extended to the subgroup analysis.

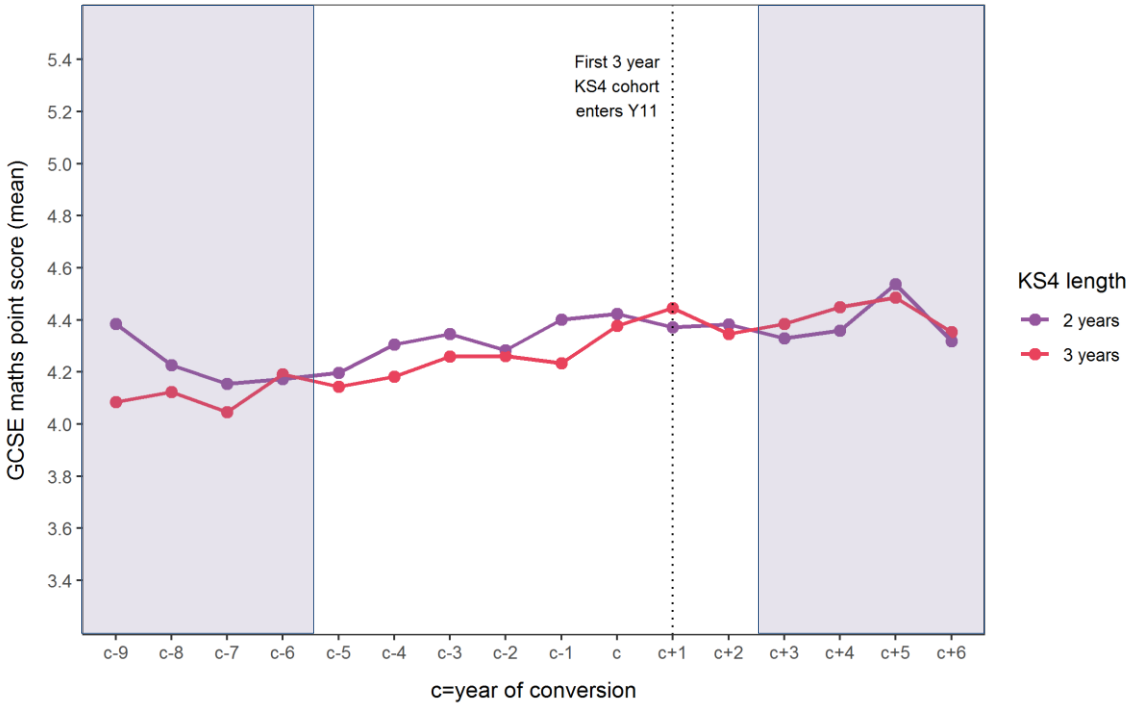
Figure 8a: Mean GCSE maths point score by length of KS4 programme and year of conversion—FSM eligible pupils only; unconditional plot

[27] The bands (quintiles) were created out of the sample used for analysis.



Source: NPD data 2007/2008 to 2018/2019.
Shaded areas indicate years with partial data.

Figure 8b: Mean GCSE maths point score by length of KS4 programme and year of conversion—FSM eligible pupils only; conditional plot

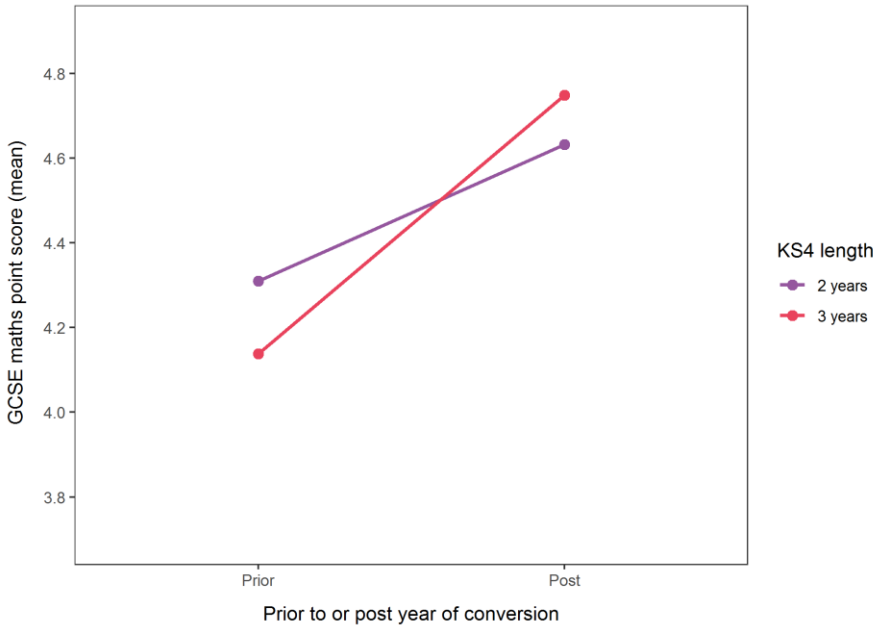


Source: NPD data 2007/2008 to 2018/2019.
Shaded areas indicate years with partial data.

Figures 9a and 9b show the mean GCSE maths point score plotted, grouped by the length of KS4 programme and whether GCSEs were taken prior to or post year of conversion for everFSM pupils only. Figure 9a shows the plot of the raw or unconditional means while Figure 9b shows the conditional plot, including school- and pupil-level covariates.

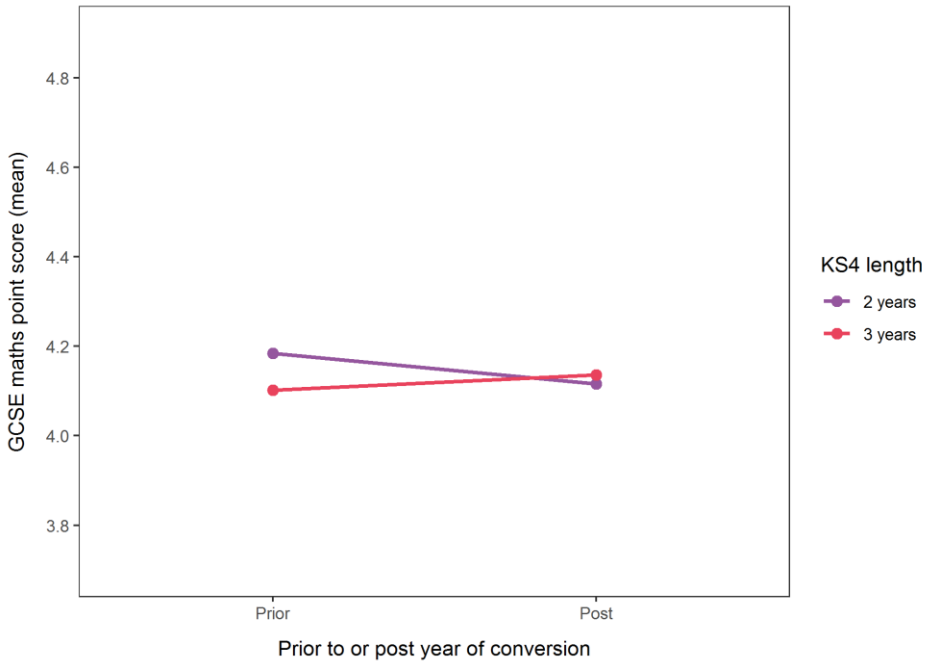
The variable *year of conversion* identifies the year in which a cohort actually sat their GCSE examinations. In the plots, time (*year of conversion*) has been collapsed into two categories: either prior to or post conversion. In terms of overall attainment, Figure 9a suggests that everFSM pupils in the three-year KS4 group improved more quickly than everFSM pupils in the two-year KS4 group. The difference in the raw mean GCSE maths point scores between 'prior to' and 'after' the year of conversion for the two-year KS4 group is 0.322 and for the three-year KS4 group is 0.611. A similar picture appears in the conditional plot (Figure 9a), which takes into account pupil and school characteristics.

Figure 9a: Mean GCSE maths point score by length of KS4 programme and whether the year is prior to or post year of conversion—everFSM pupils only; unconditional plot



Source: NPD data 2007/2008 to 2018/2019.

Figure 9b: Mean GCSE maths point score by length of KS4 programme and whether the year is prior to or post year of conversion—everFSM pupils only; conditional plot



Source: NPD data 2007/2008 to 2018/2019.

The multilevel linear regression model considering everFSM pupils only indicates a statistically significant difference-in-differences effect size estimate (0.059; 95% CI: 0.016, 0.102),[28] which is of similar size to the same estimate in the primary analysis (0.046; 95% CI: 0.032, 0.061). This estimate suggests that switching to a three-year KS4 maths programme increases GCSE maths scores for everFSM pupils, however, we recommend some caution in this result as although the everFSM pre-treatment plots indicated a similar trajectory, we have not conducted placebo tests on this subgroup and we take into account the parallel trends violation of the primary analysis.

Table 15b: Subgroup analysis—FSM eligible pupils only

Model number	Outcome	Variable of interest	Effect size (95% CI)	p-value
3	Primary model (everFSM only)	Treatment	-0.047 (-0.121, 0.027)	0.212
3	Primary model (everFSM only)	postchange	-0.039 (-0.089, 0.010)	0.120
3	Primary model (everFSM only)	Difference-in-differences estimate	0.059 (0.016, 0.102)	0.007

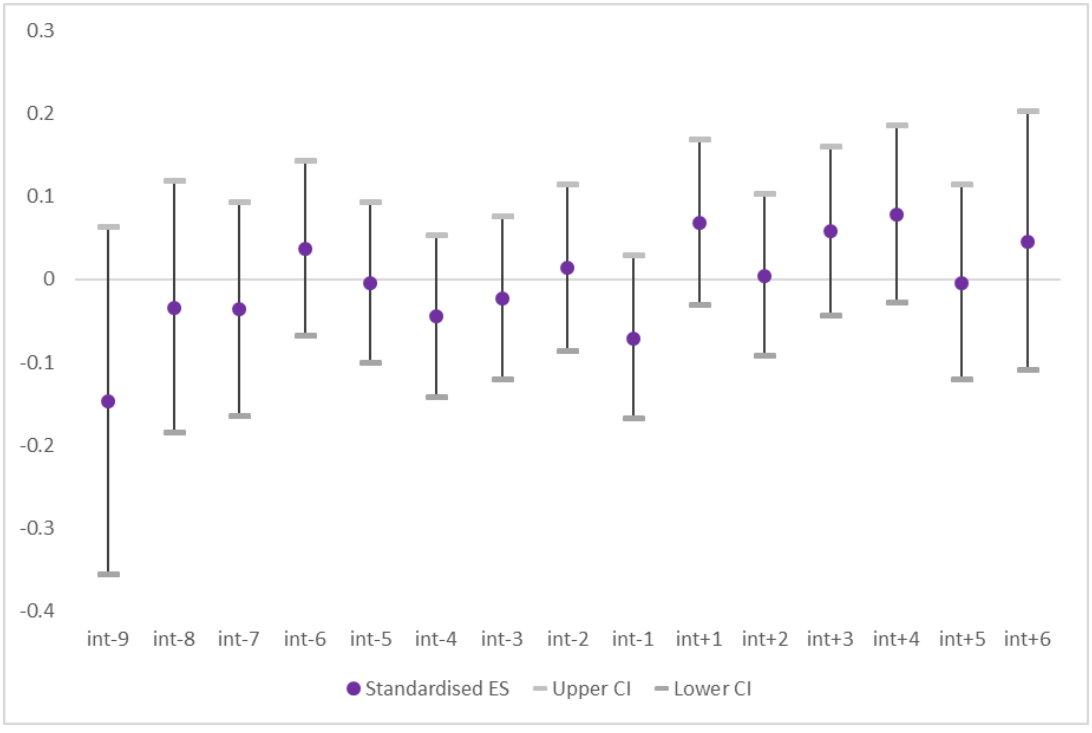
Source: NPD data 2007/2008 to 2018/2019.

Exploration of heterogeneity over time: subgroup analysis - maths, FSM pupils

We investigated whether the potential impact of a three-year KS4 on everFSM pupils varied year-on-year in the post-treatment period using the same methods as those applied above for all pupils included in the primary analysis.

Figure 10 displays the 'int' variable, which again is the interaction between treatment group (two- or three-year KS4) and year (*c*-9 to *c*+6, where *c* is the reference level).[29] While the post-treatment interaction estimates are slightly higher than in Figure 6 they tell a similar story: there is no evidence that the maths attainment gap between pupils at two- and three-year KS4 schools changes in any post-treatment year (relative to year *c*). As for Figure 6, this is likely due to changes in mean scores being relative to year *c*, rather than the entire pre-treatment period (as in Table 15b).

Figure 10: Exploration of heterogeneity over time - maths, everFSM pupils only, interaction of year (c-9 to c+6, where c is the reference level) with treatment



[28] After accounting for pupil and school level characteristics the effect size is statistically significant but small.
[29] Also see Appendix M for regression model effect sizes.

Source: NPD data 2007/2008 to 2018/2019.

Missing data analysis

As described in the Methods section, we explored the association of missingness with observable school and pupil variables with regard to the primary model. As there were no pupils missing the outcome measure, the probability that the prior attainment measure was missing (compared to observed) was modelled using a multilevel logistic model.
In total, we had 16,008 pupils with missing KS2 maths scores; 111 pupils had missing FSM and SEN information and 1,008 pupils had missing school size information. The percentage of missing is shown in Table 16. As you can see, only the prior attainment measure has more than 5% missing. The probability that the outcome measure was missing was found to be significantly associated with ethnicity, SEN, FSM, and prior attainment. Results of this analysis is found in the technical appendix (Appendix L).

Table 16: Missing data pattern

Variable	Missing N	Missing %
Outcome	0	0
Postchange	0	0
Group	0	0
Gender	0	0
Ethnicity	0	0
SEN	111	0.059
FSM (pupil level)	111	0.059
School size	1008	0.537
FSM (school level)	0	0
Region	0	0
Prior attainment	16008	8.526
Year	0	0

Source: NPD data 2007/2008 to 2018/2019.

These patterns of missing data demonstrate that the data was not missing completely at random (not MCAR).

Missing data was imputed (with chained equations, implemented using the MICE package in R: van Buuren and Groothuis-Oudshoorn, 2011) under the assumption that data was missing at random (MAR). We ran five different imputation models, each with 50 iterations. Each model included the primary outcome variable as well as covariates found in the primary model.

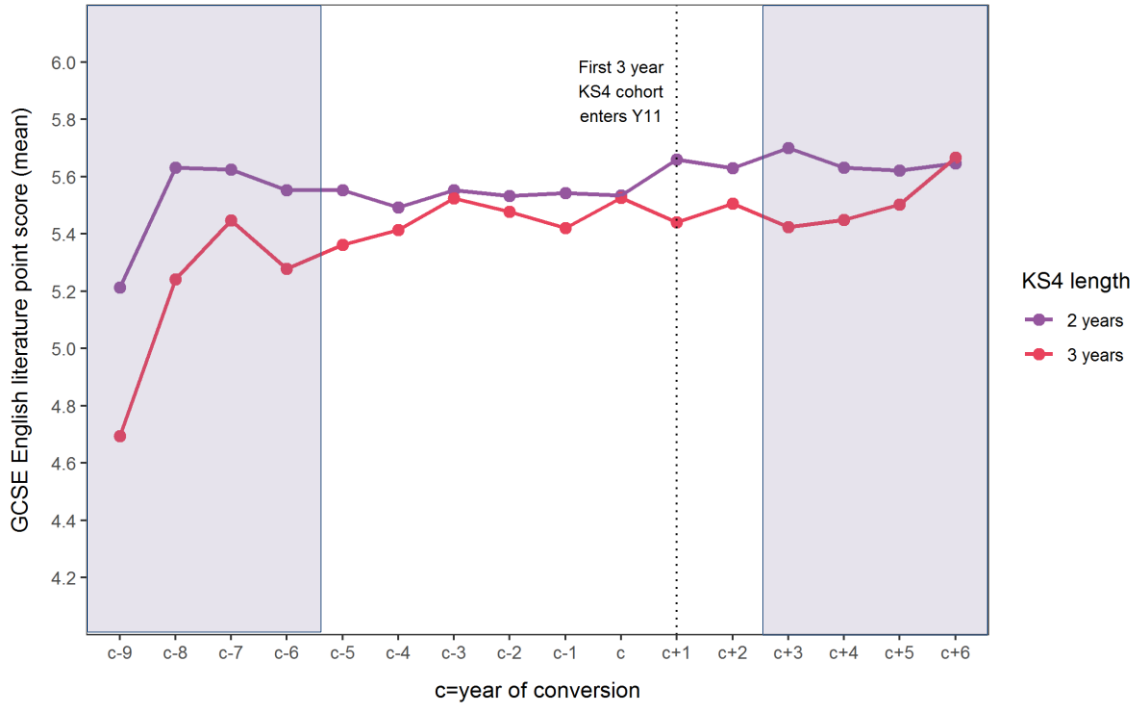
The main model was run using each of the imputed datasets. The results from the imputed datasets were pooled to give coefficients and standard errors that took account of the imputation variance. The complete data analysis gave the coefficient of being in the three-year KS4 maths group as 0.095 (0.073, 0.117). This compares to a completers model raw intervention coefficient of 0.094 (0.084, 0.104). These results from the imputed models imply that even with imputed values for the prior attainment (KS2 maths), the results were fairly consistent with the primary models and we could be certain that the completers analyses are unlikely to be biased.

Secondary outcome (1) - English literature

Again, we follow a similar structure to that of the primary analysis findings: we present charts showing the mean GCSE English literature score trajectories before presenting the findings of the model.

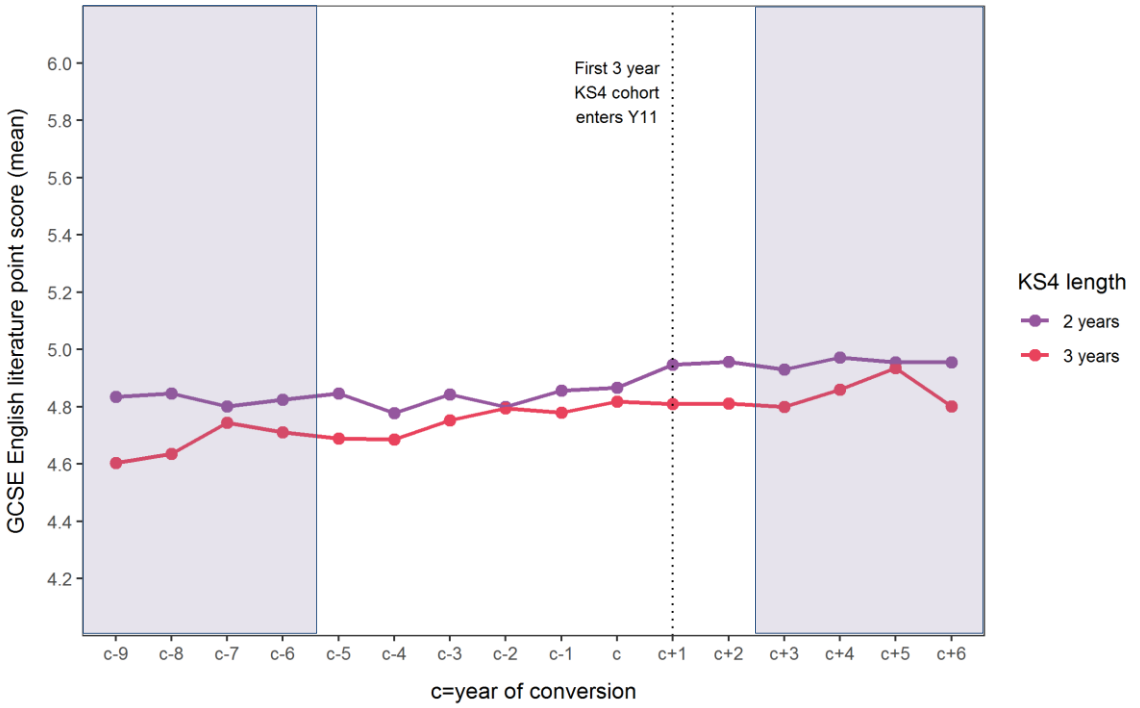
Figures 11a and 11b (unconditional and conditional plots, respectively) plot the mean GCSE English literature point score, grouped by the length of KS4 programme and year of conversion. These charts show an overall trend that pupils in the two-year KS4 group perform better than the three-year KS4 group. The key timepoints to consider are those between '*c*-5' and '*c*+2' because it is during this period that all analysed schools contribute data, indicated by the non-shaded area in the figures. Between the years '*c*-5' to '*c*+2', we can see that the gap between the two lengths of KS4 English literature fluctuates over time in both the conditional and unconditional versions of the plot but appears - upon visual inspection at least - to be more similar than the convergence observed in the plots for maths (Figures 4a and 4b). However, the apparent parallel trends violation observed for maths (the primary outcome) should also herald caution around the findings of the secondary analysis reported here.

Figure 11a: Mean GCSE English literature point score by length of KS4 programme and year of conversion—unconditional



Source: NPD data 2007/2008 to 2018/2019.
Shaded areas indicate periods with partial data.

Figure 11b: Mean GCSE English literature point score by length of KS4 programme and year of conversion—conditional

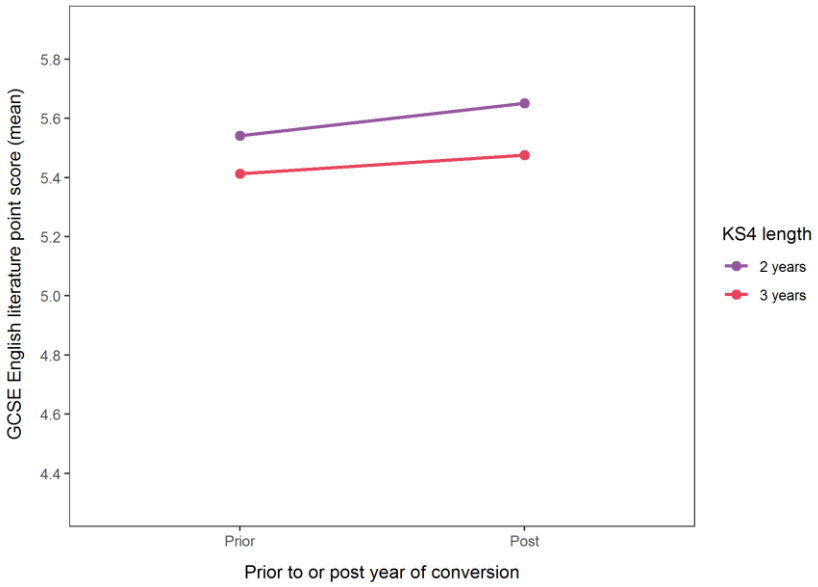


Source: NPD data 2007/2008 to 2018/2019.
Shaded areas indicate periods with partial data.

Figures 12a and 12b (unconditional and conditional plots, respectively) show plots of the mean GCSE English literature point score grouped by the length of KS4 programme and whether GCSEs were taken prior to or post year of conversion. This needs to be interpreted in the context of how the groups were selected, that is, according to the primary outcome, which was whether or not the school had a three-year KS4 for maths. The groups were not re-selected for the secondary outcome analysis but, upon checking, in almost all cases schools in the primary outcome analysis had the same length of KS4 for English literature as maths: 31 of the 32 'three-year maths' schools also had a three-year KS4 for English literature, and all of the 'two-year maths' schools had a two-year KS4 for all subjects (that is, including English literature).

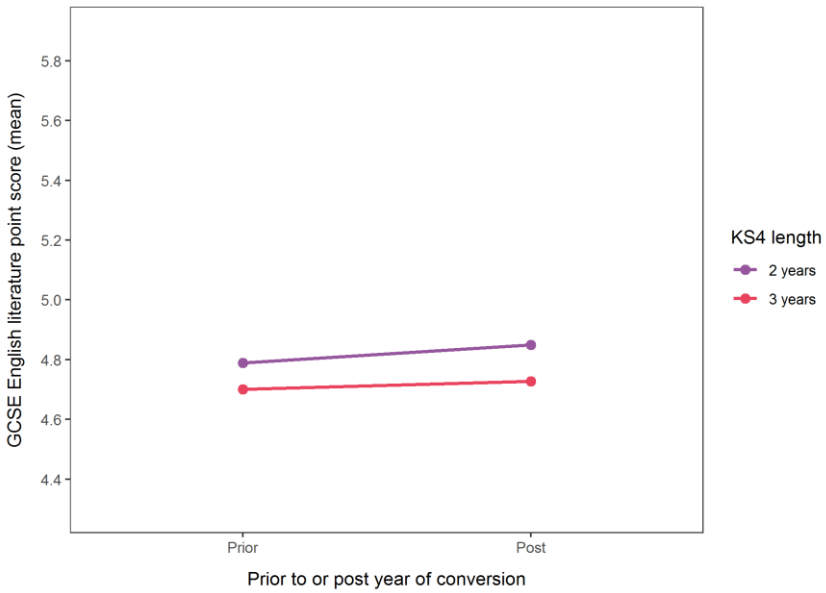
The variable *year of conversion* identifies the year in which a cohort actually sat their GCSE examinations. Here the year of conversion is collapsed into either prior to year of conversion or post year of conversion. The unconditional (Figure 12a) and conditional plots (Figure 12b) appear to show a similar picture. Overall, pupils in the two-year KS4 group appear to perform better in English literature than pupils in the three-year KS4 group. Additionally, if we consider the trend over time, we can see that in English literature the two-year KS4 group appears to be improving more over time than the three-year group.

Figure 12a: Mean GCSE English literature point score by length of KS4 programme and whether the year is prior to or post year of conversion—unconditional (raw data)



Source: NPD data 2007/2008 to 2018/2019.

Figure 12b: Mean GCSE English literature point score by length of KS4 programme and whether the year is prior to or post year of conversion—conditional plot



Source: NPD data 2007/2008 to 2018/2019.

As with the primary outcome, we ran a multilevel regression model to take into account pupil- and school-level characteristics for English literature (Table 17).

Initially, to compare pupils' GCSE English literature scores between groups pre-conversion, we consider the 'treatment' variable, which did not have a statistically significant coefficient. This indicates that there was no difference in the performance of the two-year and three-year KS4 schools in English literature prior to the policy being introduced by the latter group. However, there is a significant coefficient for the *postchange* variable indicating that, in the comparison group there were significant changes over time in two-year schools even in the absence of the policy change.[30] In the comparison schools (two-year KS4), pupils' GCSE English literature scores were higher during the post-policy-change period compared to the pre-policy period. The estimate (95% CI) for English literature was 0.042 (0.024, 0.060). So, looking at what happened in the two-year KS4 schools in the absence of any change to the length of KS4, and after controlling for pupil- and school-level characteristics, this indicates that English literature scores were improving over time in comparison (two-year) schools. This can be seen in the plots above.

The next step was to investigate whether there is a difference between the groups in terms of the degree of improvement in exam scores (English literature or achieving five A* to Cs) seen post policy change. This interaction between *postchange* and *treatment* is the main parameter of interest in these models and estimates the effect of treatment on the treated (ATT). This is shown as the 'difference-in-differences' estimate in Table 17. The negative effect size for the difference-in-differences estimate suggests that GCSE English literature scores in schools with a three-year KS4[31] decreased over time in comparison to what we would expect in the absence of the policy change (as proxied by schools that did not make this change). The effect size (95% CI) for the difference-in-differences variable for secondary outcome model 1 although significant, is very small: -0.023 (-0.040; -0.005). Although the plots above suggest the parallel trends may hold for English literature, note that we have not conducted placebo tests on this secondary outcome and we take into account the parallel trends violation of the primary analysis. Therefore, we recommend caution when interpreting this finding and we do not conclude a causal effect here due to the caveats mentioned above.

Table 17: Secondary analysis results - English literature

		Length of KS4 programme				
		Three-year KS4 group	Two-year KS4 group			
Outcome	Variable of interest	n (missing)	n (missing)	Total n (intervention; control)	Standardised effect size/odds ratio (95% CI)	p-value
GCSE English scores	Treatment	45578 (0)	114022 (0)	159600 (45578; 14022)	-0.062 (-0.141, 0.017)	0.142
	Postchange	45578 (0)	114022 (0)	159600 (45578; 14022)	0.042 (0.024, 0.060)	<0.001
	Difference-in-differences	45578 (0)	114022 (0)	159600 (45578; 14022)	-0.023 (-0.040; -0.005)	0.012

Source: NPD data 2007/2008 to 2018/2019.

[30] Postchange is a binary variable taking the value 0 if a pupils' GCSE mathematics point score was awarded prior to the year in which schools converted to a three-year KS4 programme and 1 if a pupils' GCSE mathematics point score was awarded after the year in which schools converted to a three-year KS4 programme

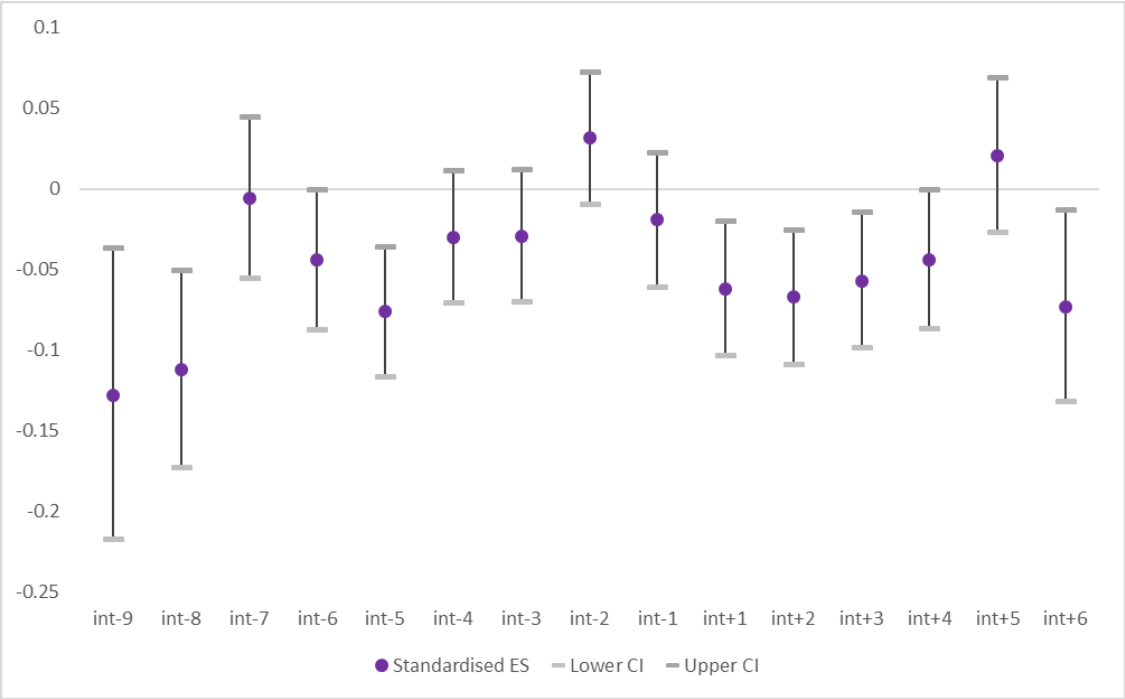
[31] Three-year mathematics programme: recall in almost all cases in the analysis schools had also converted to a longer English KS4 too.

Exploration of heterogeneity over time - English literature

As for the GCSE maths outcome, we investigated whether the impact of a three-year KS4 on GCSE English literature scores varied year-on-year in the post-treatment period using the framework described in the Methods section (Exploration of Heterogeneity Over Time).

As previously, Figure 13 displays the 'int' variable, which is the interaction between treatment group (two- or three-year KS4) and year (*c*-9 to *c*+6, where *c* is the reference level).[32] It appears that the English scores of pupils at three-year KS4 schools fall further behind those at two-year schools in the first treatment year (relative to year *c*) and that this attainment gap remains fairly constant in the years that follow. While the 'int+5' estimate may seem to upset this trend, it should be remembered that not all schools provide data in these later years making it increasingly likely that changes in the makeup of the underlying schools are conflated with a time trend. The caveat that the parallel trends assumption may well be violated, preventing a causal interpretation of these trends, also applies here as in previous sections.

Figure 13: Exploration of heterogeneity over time - English literature, interaction of year (c-9 to c+6, where c is the reference level) with treatment



Source: NPD data 2007/2008 to 2018/2019.

Secondary outcome (2) - likelihood of achieving five A* to C grades

The chart plotting the trend was only possible for the English literature outcome because the results of the model for the likelihood of achieving five A* to C grades is in the form of an odds ratio (see Table 18).

Table 18 shows the proportion of pupils obtaining five A* to C grades, or the equivalent, by length of KS4 programme.[33] As shown in the table, more pupils achieved this in the two-year (maths) KS4 group than the three-year group. There is a smaller percentage difference in the outcome for the three-year KS4 group (6%) compared to the two-year KS4 group (12%). Further analysis, taking into account pupil- and school-level characteristics, is described below.

[32] Also see Appendix M for regression model effect sizes.
[33] Again, recall that the length of KS4 is defined according to the groupings for the primary outcome and, therefore, based on the length of KS4 mathematics in the school. Some schools delivered a three-year KS4 for some, but not all, subjects (see IPE).

Table 18: Proportion of pupils obtaining five A to C grades, or equivalent, by length of KS4 programme*

Outcome		Length of KS4 programme			
		Two years		Three years	
		n/N (missing)	%	n/N (missing)	%
Obtained 5 A*–C (or equivalent)	Yes	74523/133257 (0)	56%	27234/557044 (0)	48%
	No	58734/133257 (0)	44%	29810/557044 (0)	52%

Source: NPD data 2007/2008 to 2018/2019.

As with the primary outcome, we ran a difference-in-differences model to take into account pupil- and school-level characteristics. Table 18 reports results of the difference-in-differences models for the secondary outcome of the likelihood of achieving five A* to C grades.

The model showed no statistically significant coefficient for the *treatment* variable indicating that there was no difference in the performance of the two-year KS4 schools and the three-year KS4 schools prior to the policy being introduced by the three-year KS4 schools.

However, there is a significant coefficient for the *postchange* variable[34] indicating that in the comparison group there were significant changes over time in two-year schools even in the absence of the policy change. The likelihood of obtaining five A* to C grades (or the equivalent) in comparison schools (two-year KS4) was *lower* during the post-policy-change period compared to pre-policy-change period - odds ratio of 0.743 (0.704, 0.783). So, looking at what happened in the two-year KS4 schools in the absence of any change to the length of KS4, and after controlling for pupil- and school-level characteristics, this indicates the likelihood of achieving five A* to Cs decreased during the same period.

The next step was to investigate whether there is a difference between the groups in terms of the degree of improvement in exam scores (achieving five A* to Cs) seen post policy change. This interaction between *postchange* and *treatment* is the main parameter of interest in these models and estimates the effect of treatment on the treated (ATT).This is shown as the 'difference-in-differences' estimate in Table 19.

There is no statistically significant evidence to suggest that switching to three-year KS4 maths affects the likelihood of obtaining five A* to C grades (or the equivalent); that is, the length of KS4 does not make a difference, when pupil- and school-level characteristics are controlled for. The odds ratio for secondary outcome model 2 is 0.969 (0.920, 1.022). Again, the same caveats apply as for the analyses reported above.

[34] *Postchange* is a binary variable taking the value 0 if a pupils' GCSE mathematics point score was awarded prior to the year in which schools converted to a three-year KS4 programme and 1 if a pupils' GCSE mathematics point score was awarded after the year in which schools converted to a three-year KS4 programme.

Table 19: Secondary analysis results – likelihood of achieving five A* to Cs

		Length of KS4 programme				
		Three-year KS4 group	Two-year KS4 group			
Outcome	Variable of interest	n (missing)	n (missing)	Total n (intervention; control)	Standardised effect size/odds ratio (95% CI)	p-value
Five A*–C (or equivalent)	Treatment	57044 (0)	133257 (0)	190301 (57044; 133257)	0.826 (0.613, 1.112)	0.208
	Postchange	57044 (0)	133257 (0)	190301 (57044; 133257)	0.743 (0.704, 0.783)	<0.001
	Difference-in-differences	57044 (0)	133257 (0)	190301 (57044; 133257)	0.969 (0.920, 1.022)	0.251

Source: NPD data 2007/2008 to 2018/2019.Note: The effect size for the 'difference-in differences' variable of secondary outcome model 1 is reported instead of the estimate.

Secondary outcome (3) - curriculum breadth

The section that follows describes the findings from the analysis of the third secondary outcome, investigating the breadth of the curriculum in the schools in the impact sample.

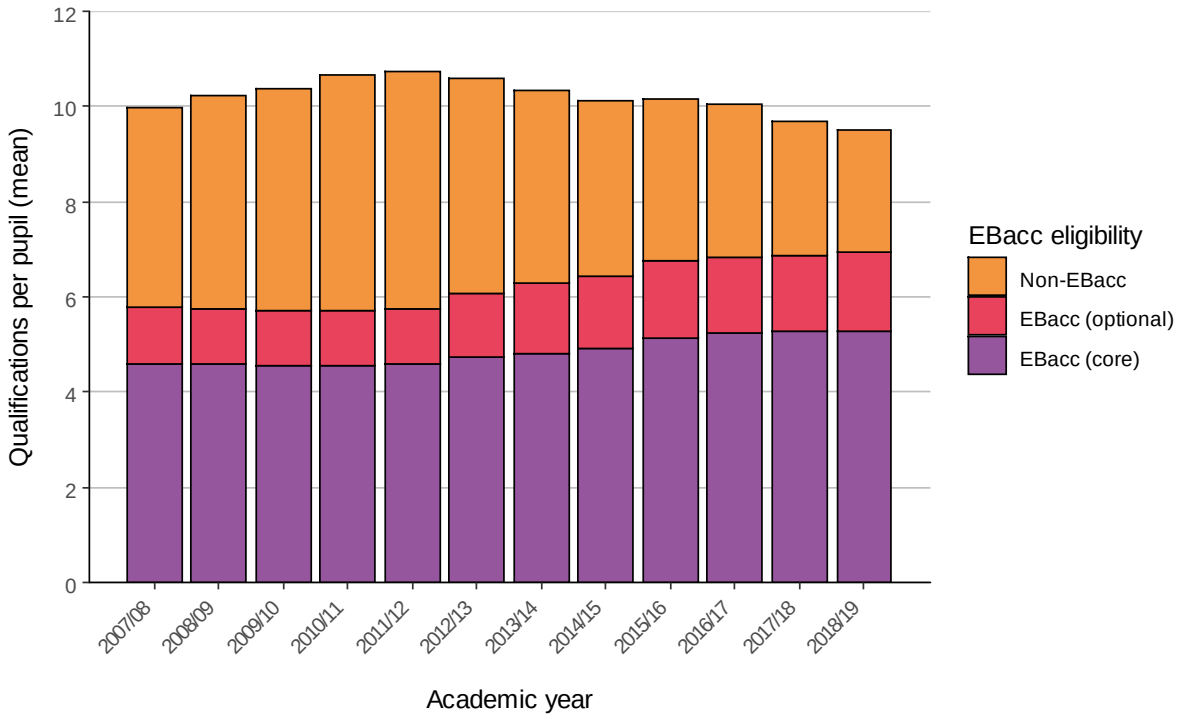
Sub-outcome (a) - mean qualifications entered per pupil

The mean number of qualifications entered per pupil—sub-outcome (a)—is plotted for all 104 schools in the impact analysis in

Figure 14. The number of qualifications per pupil generally declined from 2011/2012 (mean 10.7 onwards) reaching its lowest point in 2018/2019 (mean 9.5). This decline is due to a decrease in the number of non-EBacc qualifications entered between 2011/2012 (mean 5.0, 46.4% of qualifications taken that year) and 2018/2019 (mean 2.6, 27.1%). In

contrast, the number of EBacc-eligible qualifications ('core' and 'optional' combined) increased between 2011/2012 (mean 5.8, 53.6%) and 2018/2019 (mean 6.9, 72.9%). This reflects the pattern of uptake across all schools, reported by Gill (2018).

Figure 14: Mean number of qualifications entered per pupil for all 104 schools between 2007/2008 and 2018/2019



Source: NPD data 2007/2008 to 2018/2019.

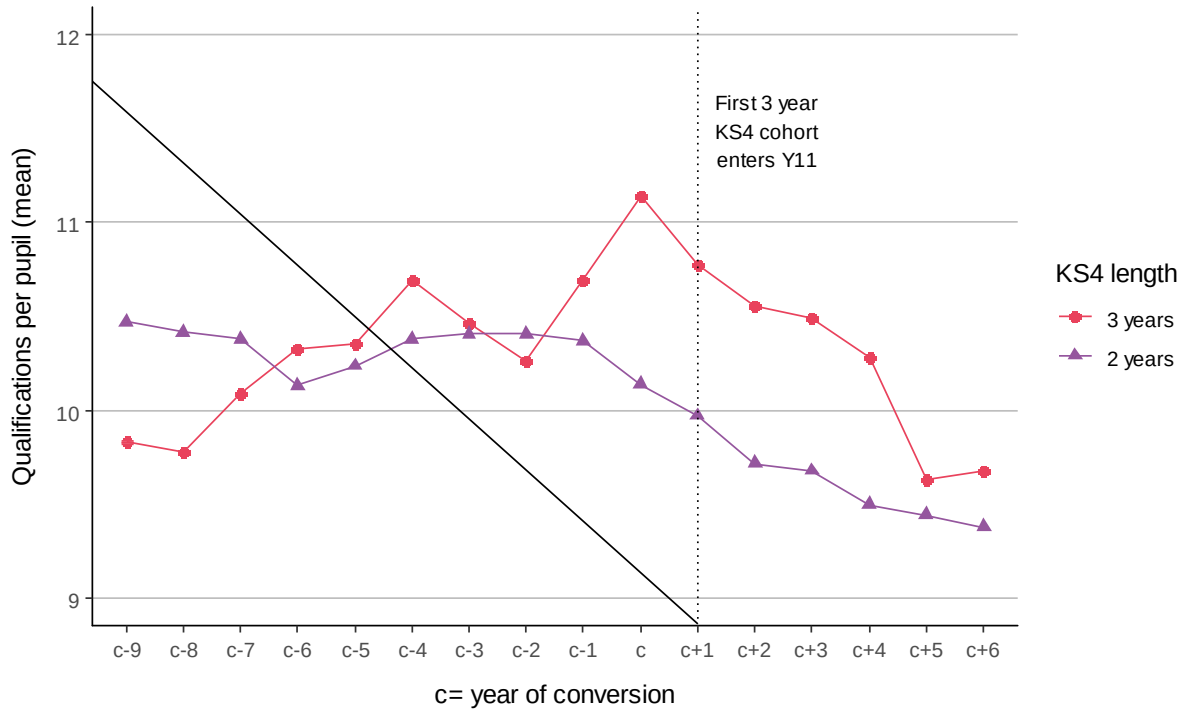
For

Figure 15 and Figure 16 the mean number of qualifications entered per pupil is again plotted, though now grouped by two- and three-year KS4 length. Recall that the x axis timeline is standardised so that '*c*+1' is the first Year 11 cohort under the three-year KS4 system (see Statistical Analysis section). It follows that if a three-year KS4 system is associated with a change in the number of qualifications entered we would expect to see a divergence in the trend of the two lines at '*c*+1' or shortly afterwards. In fact, while there does appear to be a divergence, it occurred earlier around the year '*c*', with three-year KS4 schools pulling ahead by a mean of 1.00 qualification, a gap that remained fairly constant thereafter (

Figure 15). The mean number of GCSEs entered at three-year KS4 schools is 0.6 lower at '*c*-2', then equals and overtakes two-year KS4 schools at '*c*+1', though again the year '*c*+1' does not mark a change in the trend (Figure 16). This apparent shift in the years prior to *c*=1 may be due to a change in the system at schools in preparation for the switch to a three-year KS4. The fact that the group trajectories cross before year '*c*+1' may suggest that the parallel trends assumption would not hold if a difference-in-differences analysis was performed for this outcome.[35] These comparisons should be considered in the context of the uncertainty surrounding estimates; all confidence intervals for these figures can be found in Appendix H. The confidence intervals show that at year '*c*+1' the mean number of qualifications was higher at three-year KS4 schools (95% CI: 10.3, 11.0) than at two-year KS4 schools (9.8, 10.2).

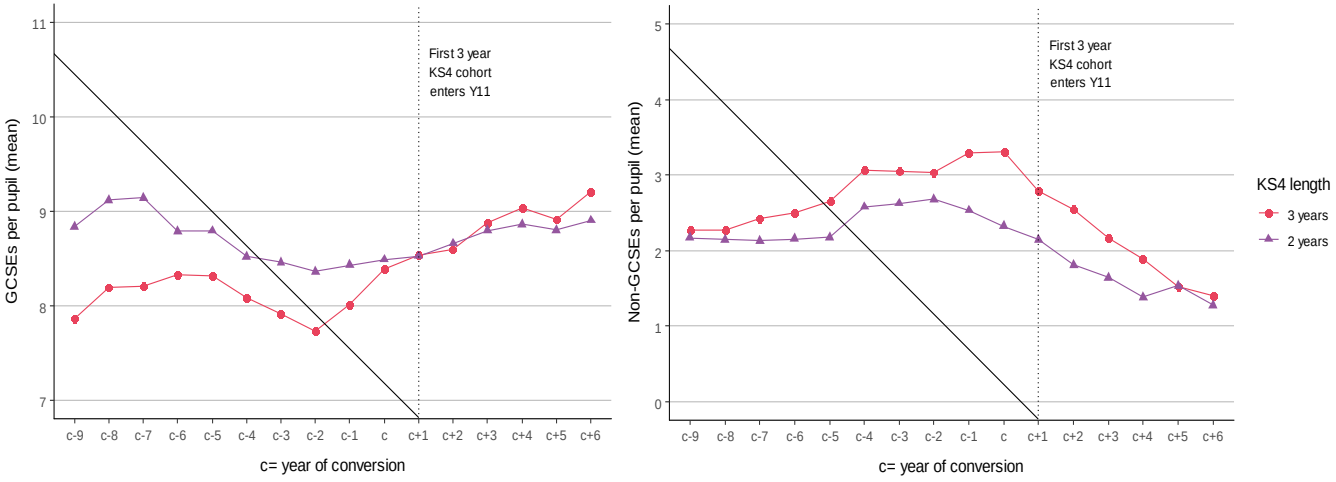
Figure 15: Mean number of qualifications entered per pupil for 32 three-year and 72 two-year KS4 schools

[35] However, note that these are simply plots of the unconditional means in the two groups and that a difference-in-differences assumption is not being formally made or tested.



Source: NPD data 2007/2008 to 2018/2019.

Figure 16: Mean number of qualifications entered per pupil for 32 three-year and 72 two-year KS4 schools—counting only GCSE (left) and non-GCSE (right) qualifications.



Source: NPD data 2007/2008 to 2018/2019.

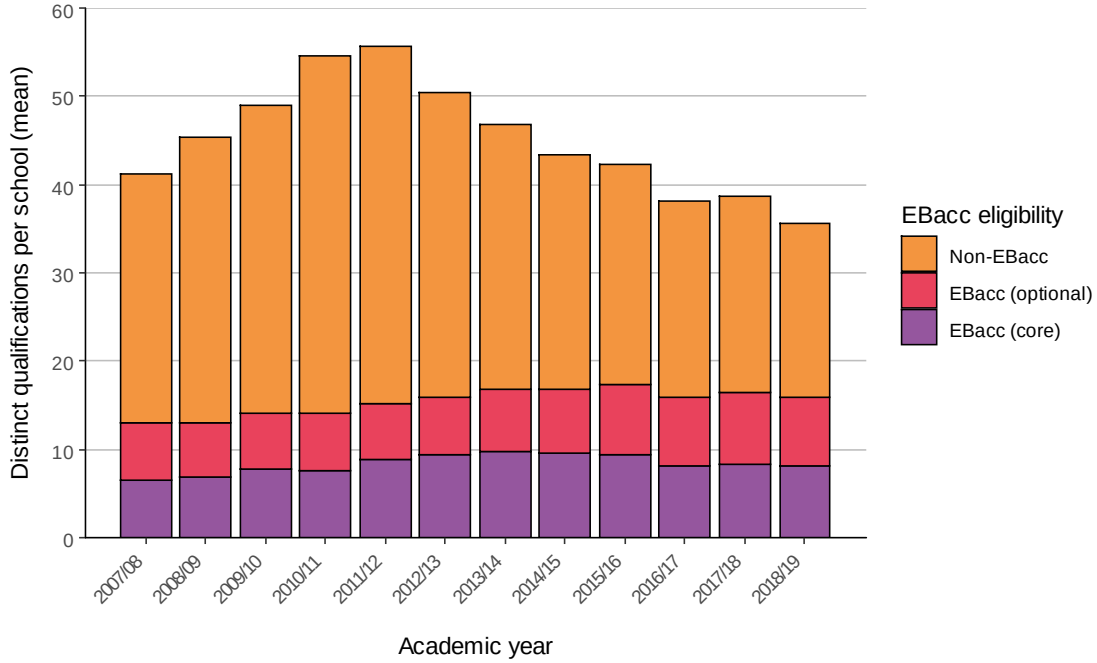
Sub-outcome (b) - mean distinct qualifications entered per school

Figure 17 displays the mean number of distinct qualifications entered per school - sub-outcome (b) - plotted for all 104 schools in the impact analysis.[36] As with sub-outcome (a), the number of distinct qualifications per school peaked in 2011/2012 (mean 55.7) but with a sharper rate of decline leading up to 2018/2019 (mean 35.2). Again, the decline was largely in non-EBacc qualifications, from a mean of 40.4 in 2011/2012 (72.6% of qualifications that year) to a mean of 19.5 (55.4%) in 2018/2019. This trend coincides with the reforms that followed the 2011 Wolf report, which meant that many vocational subjects no longer contributed to school performance tables. The number of EBacc-eligible qualifications per school increased between 2011/2012 (mean 15.3, 27.4%) and 2018/2019 (mean 15.7, 44.6%), with the number of 'optional' EBacc qualifications (for example, history) increasing by a mean of 1.6.

[36] 'Distinct' means that each qualification is counted only once, regardless of how many pupils in a cohort took it.

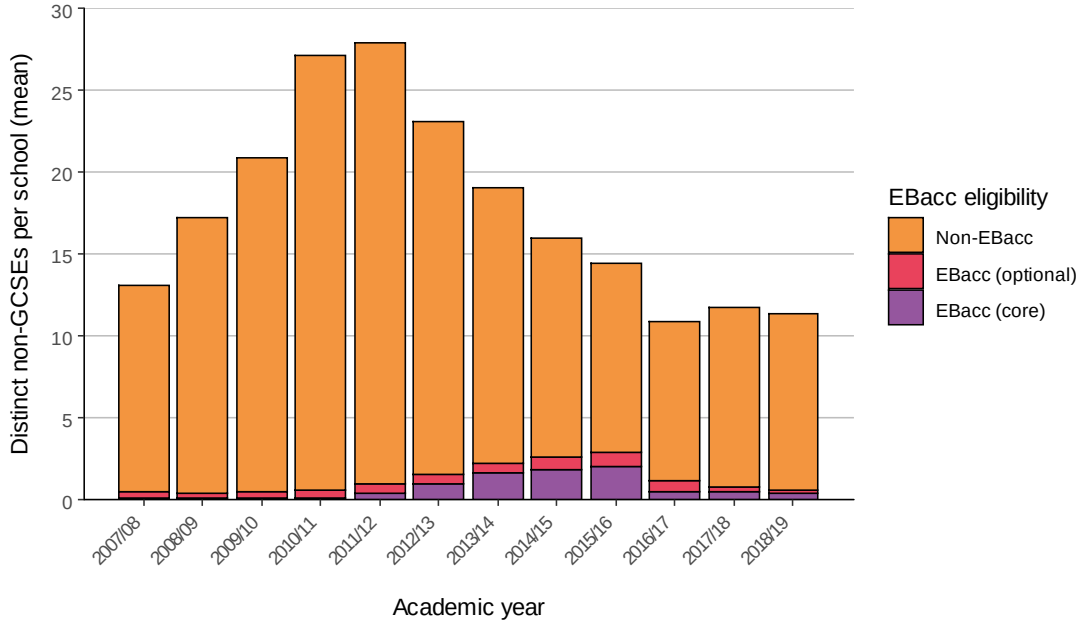
Figure 18 restricts qualifications to non-GCSEs only,[37] showing that the number of these has been particularly in decline (mean 27.9 in 2011/2012 to mean 11.3 in 2018/2019); there is of course a great deal of overlap between non-GCSEs and the non-EBacc qualifications described above.

Figure 17: Mean number of distinct qualifications entered per school for all 104 schools between 2007/2008 and 2018/2019



Source: NPD data 2007/2008 to 2018/2019.

Figure 18: Mean number of distinct non-GCSEs entered per school for all 104 schools between 2007/2008 and 2018/2019



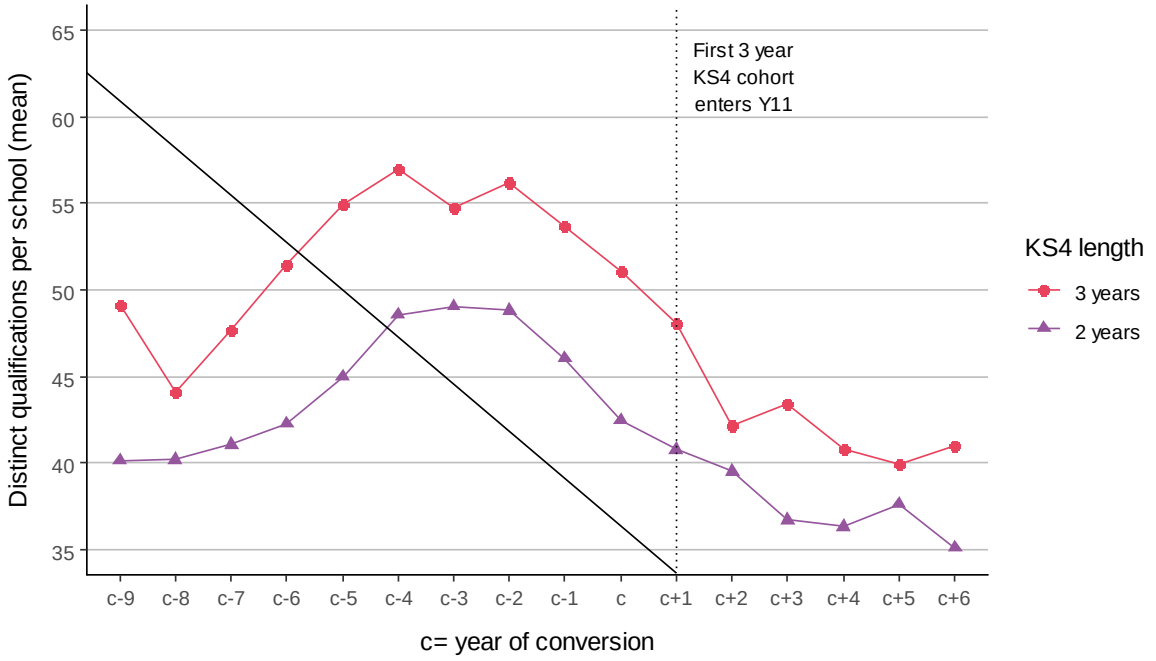
Source: NPD data 2007/2008 to 2018/2019.

The mean number of distinct qualifications per school for two- and three-year KS4 schools is displayed in Figures 19 to 21. While the three-year KS4 schools have a higher mean number of qualifications per pupil, this difference remains fairly consistent from year to year, with no divergence in the trend between the two groups. Most importantly, there is

[37] The majority are these were 'level 1/2' certificates, which were not GCSEs and could be eligible for the EBacc. These were largely phased out from 2016/2017 onwards. For an overview see: https://qips.ucas.com/qip/level-1-level-2-certificates

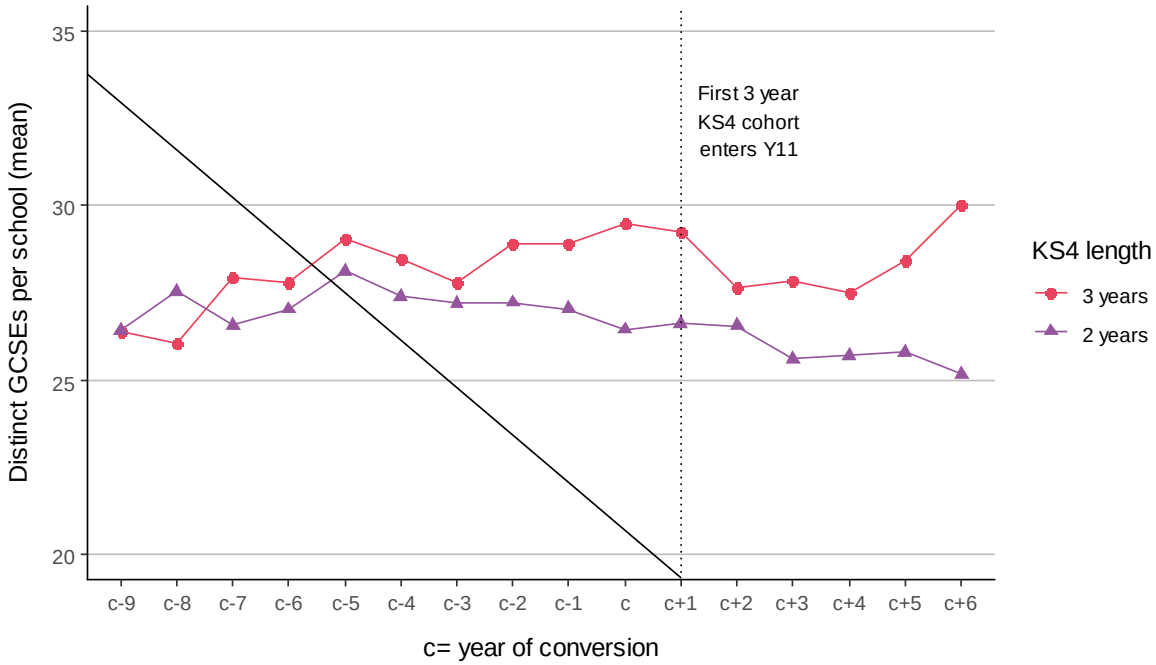
no evidence of a trend emerging at or shortly after the year '*c*+1' when the first cohort take their Year 11 exams under the three-year KS4 system. In the same way as for sub-outcome (a), these comparisons should be considered in the context of the confidence intervals surrounding estimates, which can be found in Appendix H. The confidence intervals for sub-outcome (b) are a lot wider than for sub-outcome (a): at year '*c*', 95% confidence intervals for the mean qualifications at two-year and three-year KS4 schools are (39.7, 45.2) and (45.4, 56.7) respectively. Note that the apparent upturn in number of distinct qualifications entered at '*c*+6' for three-year KS4 schools is largely due to a change in the underlying composition of the group (that is, smaller numbers of schools in the analysis) rather than a longitudinal trend, as explained in the Secondary Analysis part of the Methods section.

Figure 19: Mean number of distinct qualifications entered per school for 32 three-year and 72 two-year KS4 schools



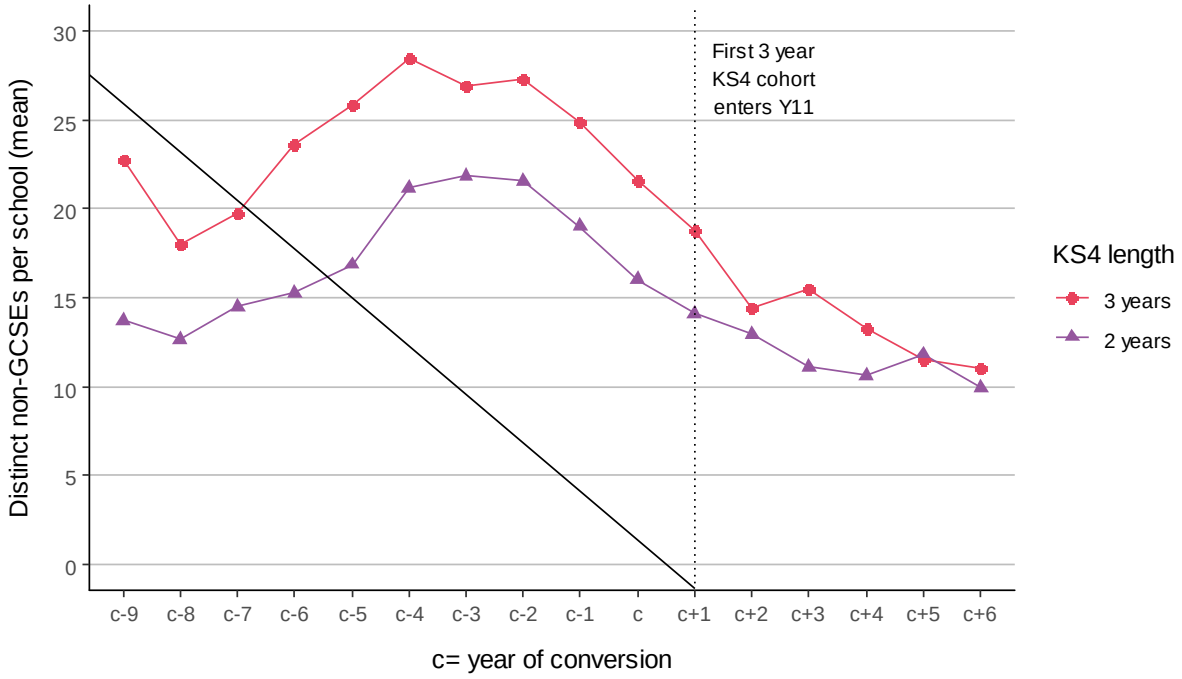
Source: NPD data 2007/2008 to 2018/2019.

Figure 20: Mean number of distinct GCSEs entered per school for 32 three-year and 72 two-year KS4 schools



Source: NPD data 2007/2008 to 2018/2019.

Figure 21: Mean number of distinct non-GCSEs entered per school for 32 three-year and 72 two-year KS4 schools



Source: NPD data 2007/2008 to 2018/2019.

Sub-outcome (c) - proportion of qualifications belonging to each subject area

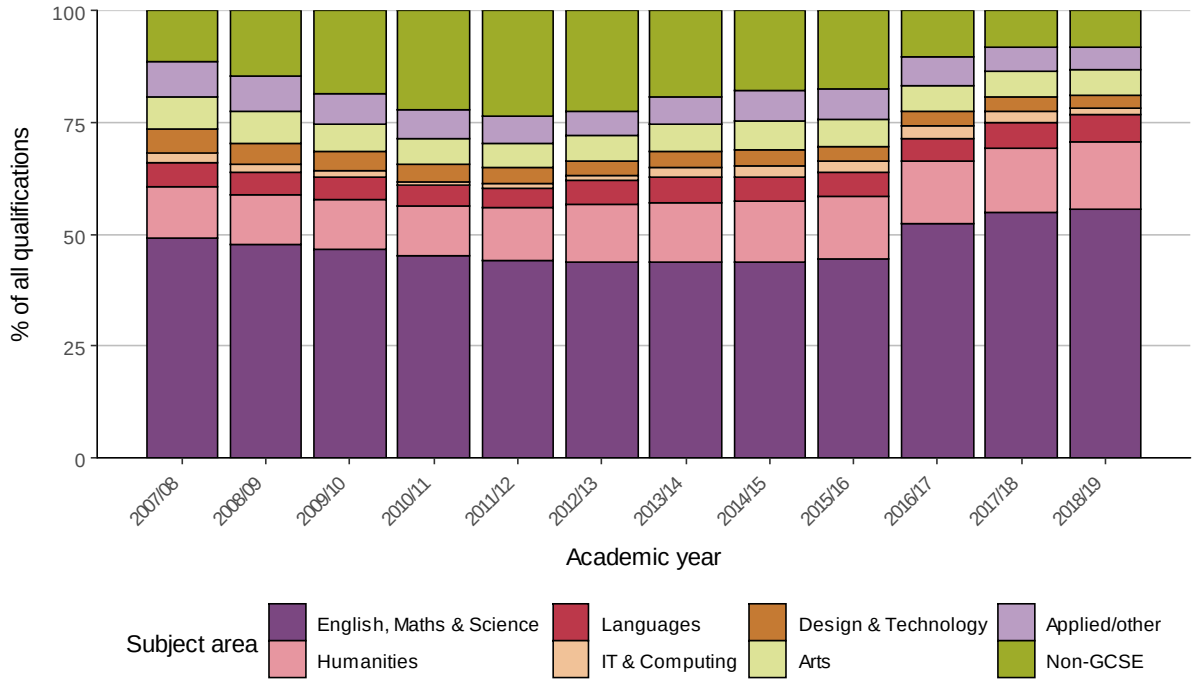
For sub-outcome (c), the proportion of all qualifications belonging to each subject area is plotted in Figure 22. The most obvious trend is the proportion of non-GCSE qualifications falling from 23.8% in 2011/2012 to 8.3% in 2018/2019. However, additional context is required: the sudden rise in the proportion of English, maths, and science qualifications and the sudden fall in the proportion of non-GCSEs in 2016/2017 is largely due to the phasing out of Level 1 and Level 2 Certificates from this point onwards. Level 1 / Level 2 Certificates were provided by awarding bodies such as AQA and were taken in the place of English, maths, and science GCSEs at some schools.[38] In Figures 22 and 23, Level 1 and Level 2 Certificates are classified as 'non-GCSE', not 'English, maths and science'.

Humanities GCSEs increased as a proportion of all qualifications year-on-year from 10.9% in 2008/2009 to 14.9% in 2018/2019. The proportion of qualifications in GCSE design and technology subjects steadily declined from 5.1% in 2007/2008 to 2.8% in 2018/2019. The trend for GCSE arts subjects is less clear as proportions fluctuated during the period analysed; the average proportion for the three most recent years of 2016/2017 to 2018/2019 (5.7%) is slightly lower than the average proportion across all years (6.1%). From 2012/2013 onwards there was no clear increase or decrease in the proportion of qualifications that were language GCSEs**,** although this proportion was higher in 2018/2019 (5.9%) than any other year in the study period (range 4.3% to 5.8%).

Figure 23 restricts sub-outcome (c) to GCSE qualifications only, making it easier to see that English, maths, science, humanities, and language qualifications combined (that is, those which often contribute to the EBacc) have increased as a proportion of GCSEs taken from 74.5% in 2007/2008 to 83.6% in 2018/2019. Over the same period the proportions of GCSEs that were taken in arts (8.2% to 6.3%), design and technology (5.7% to 3.1%), and applied or other subjects (9.1% to 5.3%) have each contracted. 'Applied/other' GCSEs include those of a more vocational nature (for example, business studies) and those that did not fit in any other category (such as sports or psychology); see technical appendix document for details.

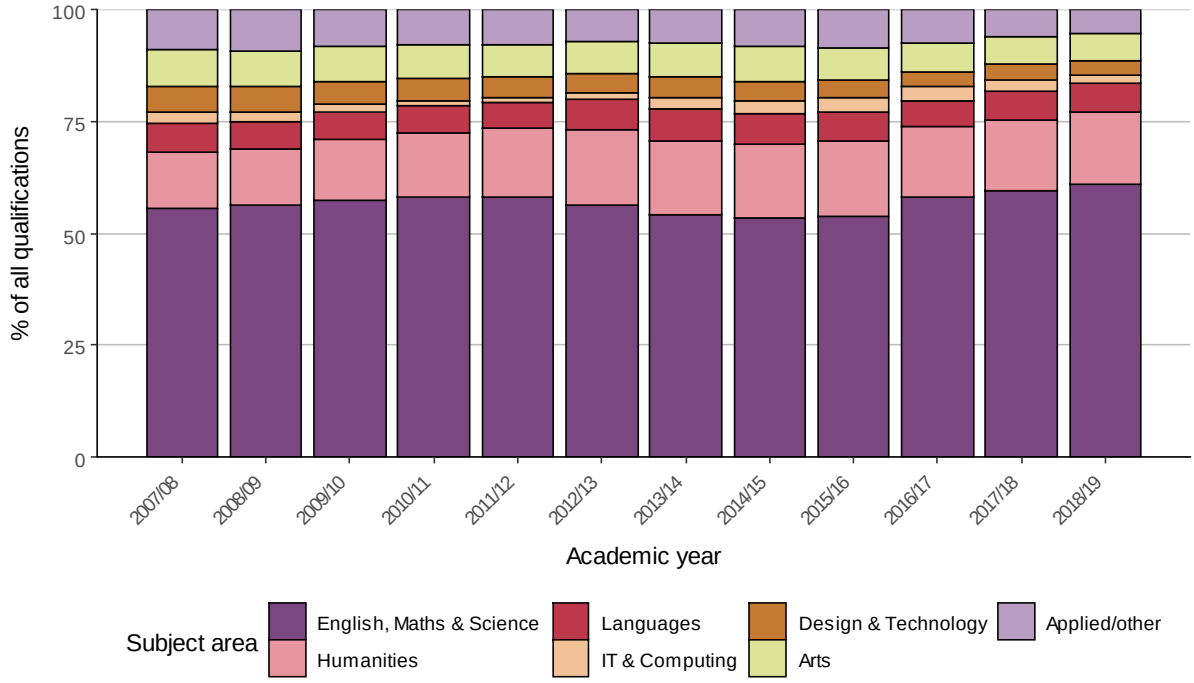
Figure 22: Overall proportion of all qualifications belonging to each subject area for all 104 schools between 2007/2008 and 2018/2019

[38] Level 1 and Level 2 qualifications were also taken in other subjects, albeit in much smaller numbers.



Source: NPD data 2007/2008 to 2018/2019.

Figure 23: Overall proportion of all GCSEs belonging to each subject area for all 104 schools between 2007/2008 and 2018/2019



Source: NPD data 2007/2008 to 2018/2019.

It is also possible to observe the shift over time in the proportion of subjects belonging to each subject area for two- and three-year KS4 schools, as shown in Figures 24a to 24c. Focusing on the time frame between *c*-2 and *c*+2 (Figure 24b), it can be seen that two-year KS4 schools generally had a higher proportion of English, maths and science subjects but there is no shift observed between the years '*c*' and '*c*+1'. There is therefore no evidence that the balance of subjects across different areas shifted as an immediate consequence of the introduction of a three-year KS4. While the proportion of subjects in each area seems to become more similar for two- and three-year KS4 schools between '*c*+3' and '*c*+6', it should again be noted that the underlying schools change from '*c*+3' onward, so apparent trends may be due to this.

Figure 24a: Overall proportion of all qualifications belonging to each subject area for two- and three-year KS4 schools between years c-7 and c-3.

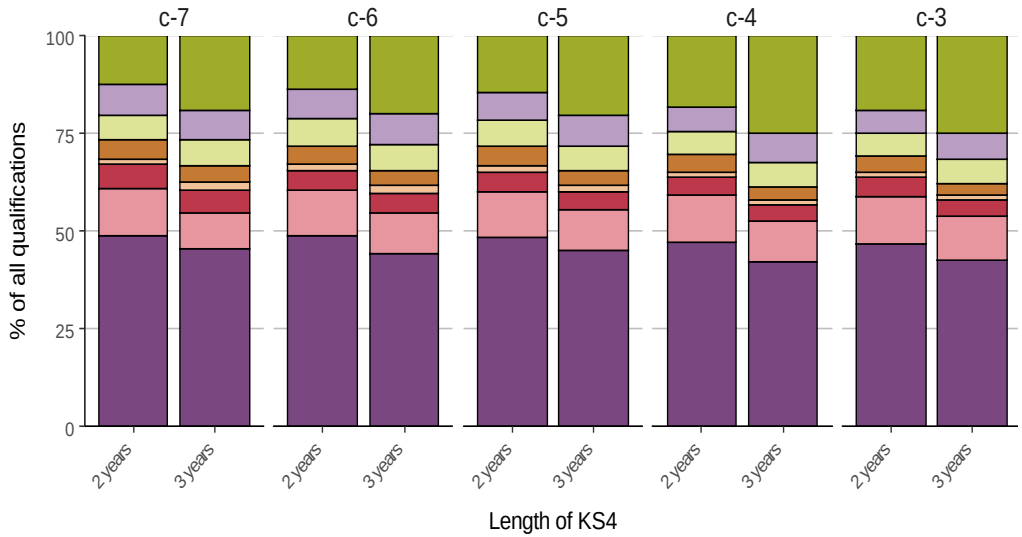


Figure 24b: Overall proportion of all qualifications belonging to each subject area for two- and three-year KS4 schools between years c-2 and c+2.

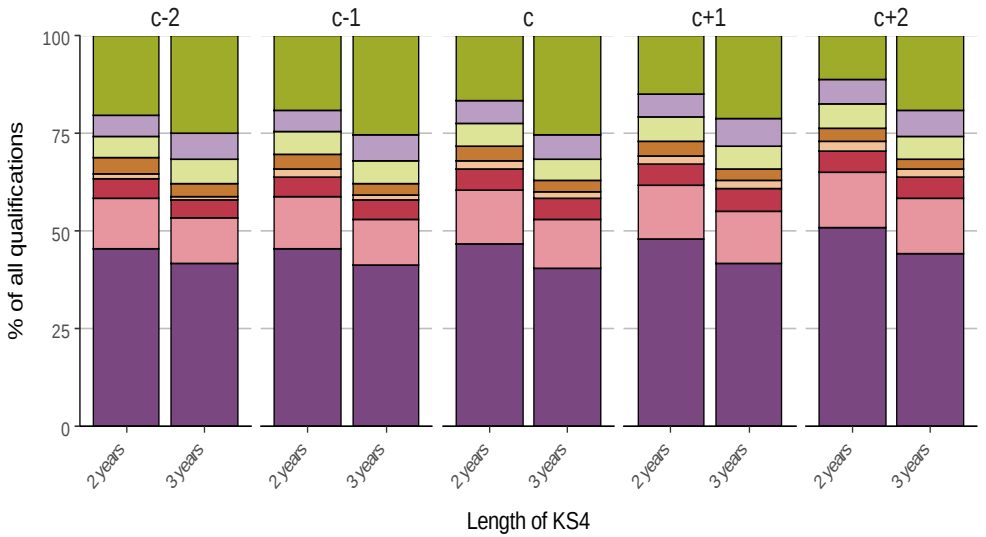
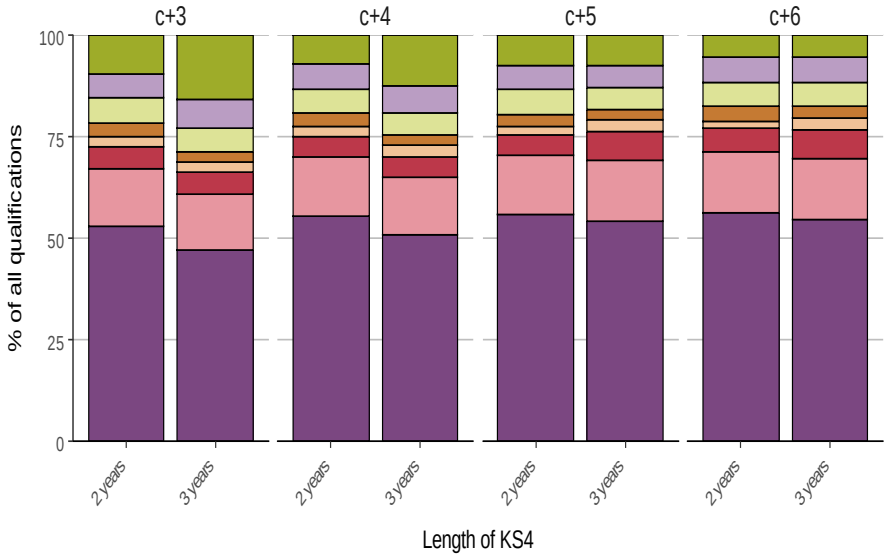
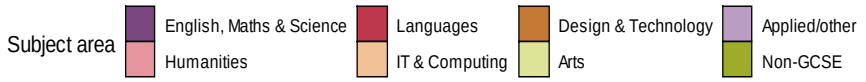


Figure 24c: Overall proportion of all qualifications belonging to each subject area for two- and three-year KS4 schools between years c+3 and c+6.



Legend:



Source: NPD data 2007/2008 to 2018/2019.

Sub-outcome (d) - proportion of pupils taking GCSEs in particular subjects

The proportion of pupils taking GCSEs in particular subjects - sub-outcome (d) - is displayed in a series of plots in Appendix F. Although uptake of languages has fluctuated, overall it had increased slightly. A decrease in French and German appears to have been offset by increased uptake of Spanish. All of the arts subjects plotted had decreased, as had design and technology subjects. In contrast, as might be expected based on the findings reported above, take-up of history and geography GCSEs had increased, while religious studies GCSE had declined.

The uptake of each GCSE was also plotted for two- and three-year KS4 schools in order to investigate whether there was a divergence in the trend between different KS4 lengths at year '*c*+1'. While for some subjects there appeared to be differing trends between two- and three-year KS4 schools over the entire study period, there was no indication of this occurring at year '*c*+1'. These results can also be found in Appendix F.

Implementation and process evaluation results

In this section we summarise the findings from the survey and the interviews. The IPE intended to answer the following research questions:

IPE_RQ1 What are the perceived impacts of having a two- or three-year programme at KS4 on pupils (in particular, but not exclusively, related to wellbeing and ability to manage workload)?

IPE_RQ2 What are the softer impacts of having a two- or three-year programme at KS4 on teachers and their lesson planning?

IPE_RQ3 What are the reasons for operating different lengths of KS3 or KS4 (for example, pupil outcomes/accountability measures/resourcing factors)? What were the factors that influenced the decision to change the length of KS3 or KS4?

IPE_RQ4 What form do different lengths of KS3 or KS4 take? What variation occurs within groups?

IPE_RQ5 What strategies and practices are used to support high-quality implementation of different lengths of KS3 or KS4?

IPE_RQ6 Which factors affect the breadth of subjects that schools offer for study at KS4 and at KS3?

The survey findings reported here include all 405 schools that responded to the survey, not only those included in the impact analysis.

The breakdown of key school characteristics of the interviewees (37 interviews with 40 individuals) is shown in Table 20.

Table 20: The characteristic composition of the schools that participated in telephone interviews

Characteristic		Number of schools
KS4 model	All pupils start formal KS4 study of some subjects in Year 9 (and the rest of their subjects in Year 10)	9
	All pupils start formal KS4 study of all subjects in Year 10	14
	All pupils start formal KS4 study of all subjects in Year 9	13
	Some pupils start formal KS4 study in Year 9 and some pupils start formal KS4 study in Year 10	1
School Type	Academy	24
	Maintained	13
Region	North	11
	Midlands	11
	South	15

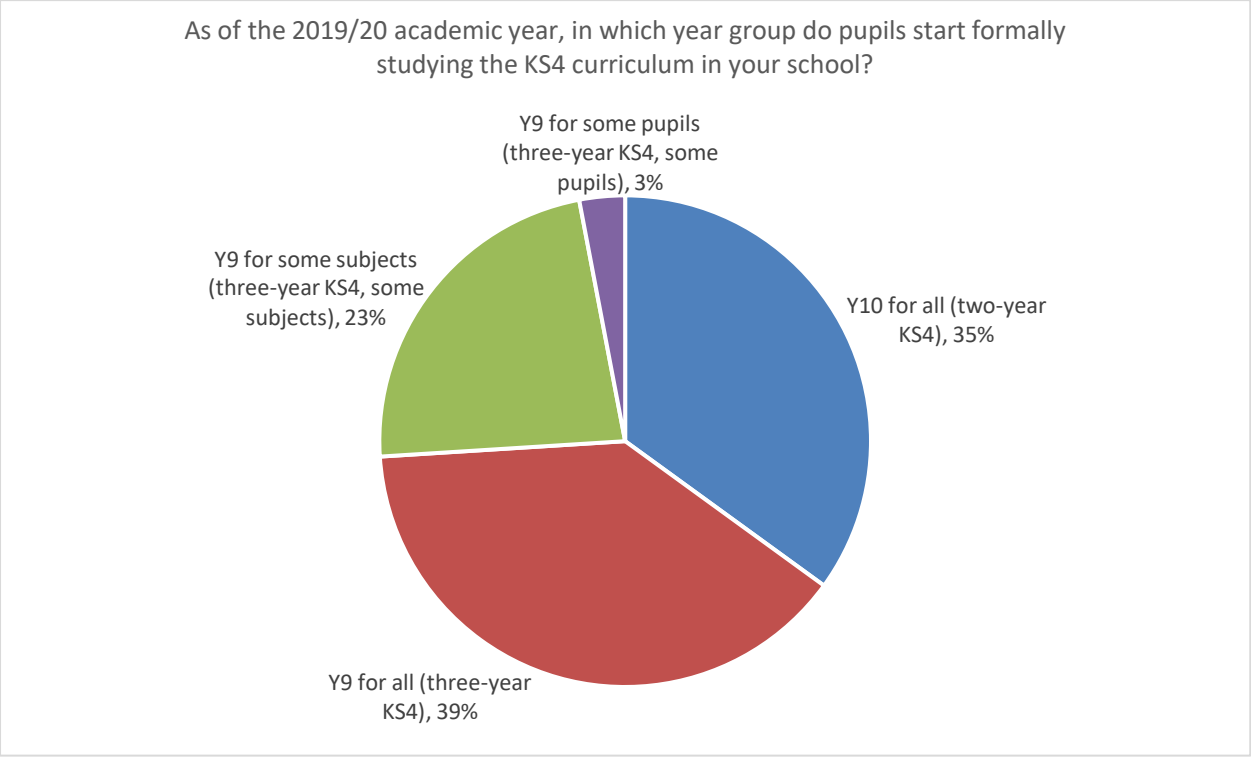
Source: NFER interviews with a sample of publicly funded mainstream secondary schools, 2020.

This section will cover the models used by schools, their reasons for doing so, and the perceived impacts on pupils, staff, and schools.

Models of KS3 and KS4 provision

Almost two thirds of schools that responded to the survey (65%) said they had a three-year KS4 for at least some subjects/pupils (Figure 25). This is only slightly higher than other similar research reported in the introduction (for example, IFF Research, 2018; NFER, 2019), however, as the survey sample was not representative of all schools, it is possible that schools that were more confident about their model responded to the survey. Schools following the national curriculum recommendation of a two-year KS4 were in the minority, with only 35% of responding schools indicating that the KS4 curriculum in their school started in Year 10. The reasons for operating different models are discussed in the next section.

Figure 25: Prevalence of KS4 models in all responding schools (%)



N = 405 schools.
Source: NFER survey of publicly funded mainstream secondary schools, 2020.

As shown in Figure 25, 23% of schools said that their pupils started studying the KS4 curriculum for some (but not all) subjects in Year 9. Table 21 shows this broken down by subject, with science and maths the most common to start in Year 9 (75% and 53% respectively). Respondents were also able to specify the 'other' subjects: responses given spanned a full range of subjects, including religious studies, IT/computing, drama, physical education and dance, and business studies.

Table 21: Schools starting KS4 in some subjects: subjects started in Year 9

Subject	Yes %	No %	No response %
English	31	65	4
Maths	53	43	4
Science	75	21	4
Humanities	29	66	4
Modern foreign languages	26	69	4
Art	21	75	4
Music	24	72	4
Technology	22	74	4
Other(s)	26	69	4

N = 95 schools.
Source: NFER survey of publicly-funded mainstream secondary schools, 2020.

Schools that started studying KS4 in Year 9 tended to start teaching the KS4 curriculum at the start of the academic year, in the autumn term, indicating a full (rather than partial) shift from a two- to a three-year KS4 (Table 22).

Table 22: Term of Year 9 in which pupils start studying the KS4 curriculum

Which term of Year 9 does KS4 start?	%
Autumn term	76
Spring term	10
Summer term	8
Not sure	3
No response	4

N = 263 schools.
A single response item.
Responses are limited to schools that began KS4 in Year 9 for at least some pupils or subjects.
Source: NFER survey of publicly funded mainstream secondary schools, 2020.

Reasons for using different lengths of KS3 or KS4

The survey respondents were able to enter free text responses to an open question about their reasons for their curriculum model. The responses were coded to take up to three reasons per school. The reasons behind different lengths of KS3 and KS4 were also explored further in the interviews (described below).

The most common reasons given were split by the type of model used by the school; none of the main reasons given overlapped by model type (Table 23). There were some similarities, for example, maintaining or maximising curriculum breadth was mentioned by both two- and three-year KS4 schools, but, demonstrating differences in focus, in relation to different key stages: two-year KS4 schools emphasised its importance at KS3, and three-year KS4 schools at KS4.

Table 23: Most common reasons for their KS3 or KS4 model given by schools responding to the survey

Two-year KS4 - start in Year 10 (n = 123)	Three-year KS4 - start in Year 9 for some or all subjects/pupils (n = 198)
Maintain/maximise the curriculum breadth/depth at KS3 (80 schools; 65%)	Additional time to study KS4 curriculum content (102 schools; 52%)
Pupil readiness (30 schools; 24%)	To improve pupil motivation/engagement in KS3 (34 schools; 17%)
The importance of a three-year KS3 as preparation for KS4 (30 schools; 24%)	Maintain/maximise the curriculum breadth/depth at KS4 (27 schools; 14%)
KS4 courses are designed to be delivered over two years (23 schools; 19%)	To improve pupil outcomes at KS4 (26 schools; 13%)
Other (15 schools; 12%)	To meet/accommodate the needs of pupils (24 schools; 12%)

Source: NFER survey of publicly funded mainstream secondary schools, 2020.
A single open response item that could be coded up to three times. Other codes with lower frequencies are not presented due to low cell counts.
Percentages are the proportion of eligible respondents within the column that received each code.
Responses are limited to respondents that completed part two of the survey (N = 321).

Schools with a two-year KS4 model reported using their model because of the importance of a strong curriculum and breadth of experience at KS3 when delivered over three years:

> *'We like to keep the depth and breadth of KS3 for as long as possible. We also feel the students are not mature enough to choose options in Year 8' (two-year KS4, survey respondent).*

> *'We have always wanted to ensure we offer a breadth of curriculum both at KS3 and KS4. We think that pupils benefit from a wider curriculum at KS3 for three years - this enables pupils to explore a wide range of subjects before specialising at KS4. We have continued to offer three MFL [modern foreign languages] at KS4 when other schools have narrowed this. We have maintained coverage of the arts and design too at KS3 and KS4. We also don't feel that pupils are ready to choose their options before Year 9' (two-year KS4, survey respondent).*

In contrast, schools with a form of three-year KS4 (that is, those that started KS4 for at least some pupils and subjects in Year 9) mentioned that they felt the additional time was needed to study the curriculum content in KS4, and that curriculum breadth and depth at KS4 was the focus:

> *'We are [a] secondary modern school in a highly selective area and our students come to us significantly behind their peers in maths and reading, with literacy scores below national as well. We have put additional time into the curriculum for English, maths, and science and this obviously reduces down the time for other subjects. To allow students to continue to study four options and have the additional time for English and maths, we need to reduce the time allocation per two-week block and extend' (three-year KS4, all pupils, all subjects, survey respondent).*

> *'More time was given to KS4 and the more difficult GCSE and BTEC courses. Starting the GCSE work in Year 9 meant students received an extra 282 hours of lessons/guided learning hours on their four option choices; 400 hours more lessons/guided learning hours on each core subject. Students would be offered four option choices at GCSE rather than the three choices that were available to them under the existing curriculum plan' (three-year KS4, all pupils all subjects, survey respondent).*

Although appearing to come from different standpoints, both types of schools emphasised the needs of pupils in different ways. Two-year KS4 schools were particularly concerned with the 'readiness' of pupils to make the options decision, choose their pathway, and to deal with the curriculum content, whereas three-year schools felt this was less of an issue in their experience. As noted above, the three-year schools generally reported that their pupils needed more time to study the content and this was the overriding motivation in using a three-year KS4 model. In contrast, other schools - including one with a three-year KS4 but only for 'core' subjects - felt that Year 8 was too soon to decide which subjects to discontinue, as the following examples illustrate:

> *'We have always believed in a broad and balanced curriculum. Selecting options in Year 8 seems too early to be dropping subjects. Year 9 is used to reinforce the learning from Year 7 and 8 and to deepen the learning required for GCSE' (two-year KS4, survey respondent).*

> *'We didn't think that it was fair to decide [about optional subjects] at the end of Year 8 when they'd only had two years of a particular subject. With greater maturity and experience of the subject, it means that they're making a more informed choice and [that can] impact on [a pupil's] pathway from GCSE to A-level to possibly degree level' (three-year KS4, all pupils, some subjects, interviewee).*

Several of the senior leaders from three-year KS4 schools described having high proportions of disadvantaged intakes. They explained that, as part of their strategy to improve social mobility, their school had increased the time for pupils to study their KS4 qualifications. Several of these schools also emphasised that by extending KS4 into three years, it meant pupils were able to take more subjects, thereby maintaining greater curriculum breadth in KS4 than would be possible in a two-year KS4. Two different schools with three-year KS4 models described their reasoning as follows:

> *'We serve a community dominated by white disadvantaged students (65% +). The vast majority of students enter school with below average literacy, below average numeracy, very low levels of cultural capital, very poor levels of self-motivation. The three-year KS4 has enabled us to complete the GCSE specifications with sufficient time to cover the content and leave time in Year 11 to revise. Outcomes have ensured good social mobility as most students achieve the grades for their next steps' (three-year KS4, all pupils, all subjects, survey respondent).*

> *'Our school serves a community with high deprivation and many of our pupils do not have digital devices, books, support for learning at home. Many have a home language other than English. We felt that to spend 18 months on the most important qualifications of a pupil's life was not enough. We have an academic curriculum which has been kept broad so all do double English, maths, double science and full course RE. Almost all do MFL and almost all choose history or geography' (three-year KS4, all pupils, all subjects, survey respondent).*

Schools delivering a three-year KS4 for some subjects particularly mentioned the need for additional time to cover the content for those subjects - most commonly science, maths and MFL. A typical comment - in this case about maths - from an interviewee illustrates this point: 'The content for the new GCSE was unmanageable and there was too much to get through in the two-year period'. This was a key theme that emerged in the interviews. School leaders in schools with a three-year KS4 explained that the introduction of the reformed GSCEs in 2015, or the schools' experiences of

delivering them soon after, prompted a review of curriculum arrangements. As a result, some schools opted to move to a three-year KS4 for some or all subjects due to the new changes to the content and structure of the reformed GCSEs. Furthermore, the 'huge shift in the number of exams which students are expected to sit at the end of Year 11' (three-year KS4, all subjects) was another factor contributing to the move to a longer KS4 for some schools. One senior leader, whose school had recently moved back from offering a three-year KS4 to a two-year KS4 explained that at the time the new GCSES were introduced, 'subject leaders were very concerned that they would not be able to deliver the required content owing to the new GCSEs being more intensive'. However, once their staff had become more familiar with the new GCSEs, they had found that 'the content in it is deliverable [in two years] if you have a carefully constructed framework for the curriculum'. Another interviewee made a similar point, arguing that KS4 content could be made more manageable by ensuring the necessary foundations were provided at KS3:

> *'A long time ago, science was three year and then changed to two year. The core content is the same and they [the science department] simply deepened KS3 to build the foundations for supporting this KS4 content' (three-year KS4 for some subjects, interview).*

(Curriculum breadth and depth is discussed in further detail in a section below.)

Senior leaders in some three-year KS4 schools described introducing the KS4 content for core subjects (maths, science, and in some cases humanities or MFL) part way through Year 9 in such a way that the pupils were unlikely to notice. As the approach was usually used for core subjects it was unrelated to making options choices, which still occurred for the remaining subjects in Year 9 to start KS4 study in Year 10. This hybrid approach with differing lengths of key stages for different subjects in some cases was designed to avoid asking pupils to make their options choices in Year 8 - as highlighted above, one of the key reasons given against a three-year KS4 by schools with a two-year model. Schools with mixed lengths of key stages for different subjects also felt that it gave more flexibility in terms of a broad curriculum at KS3, and more time for KS4 content for core subjects.

Some interviewees - and this included those in schools with both two- and three-year KS4 models - emphasised they did not view the curriculum as KS3 distinct from KS4 but rather as a five-year continuum progressing up to assessments at the end of Year 11: 'It's less about Key Stage 3, Key Stage 4 models and more about how well you are sequencing the curriculum' (three-year KS4 school).

Several interviewees in schools that offered a three-year KS4 reported that their staff had, prior to converting to a longer KS4, become aware of weaknesses in KS3 coverage in terms of the coherence of the curriculum, depth, and sequencing. One senior leader, whose school had moved from offering a two- to a three-year KS4, reported that there had been a 'disconnect between subjects' with, for example, tangents being taught in Year 10 in maths but in Year 9 in science. As a result, staff had worked to improve curriculum coherence, to increase the pace of teaching, and to avoid repetition with the primary curriculum at KS3. Indeed, most interviewees working in schools that introduced KS4 content and concepts in Year 9 did not report any issues for their pupils in accessing the curriculum. The onus was seen to be on the teachers to think about the planning and sequencing at the appropriate level:

> *'You have to be clever with your scheme of work as to what you introduce in Year 9 and what you might leave until Year 10 and 11, but I think it's perfectly possible for them to get to grips with the subject' (three-year KS4 for some subjects, interviewee).*

A minority of schools felt that although on the whole their Year 9 pupils were able to manage with the content, there were some challenges, such as the level of the assessment questions (designed for pupils at the end of Year 11) being presented to students much earlier - this required further adaptation from the teachers. Some interviewees reported identifying a general drop in pupil performance and/or engagement in Year 9 and looked to curriculum reform as a way to address the problem. For example, the senior leader of a school that had moved from offering a three-year KS4 to a two-year KS4 reported that some Year 9 pupils were 'switching off' and 'treading water' since that change. This was something they had not experienced when the school was operating a three-year KS4 model, although the interviewee could not rule out that this was associated with the pandemic and pupils feeling fatigued following a long period of absence from the classroom. Others reported wanting to introduce curriculum reforms to better meet the needs of particular student groups, including the 'middle ability' and disadvantaged students.

Making changes to the curriculum

In terms of how schools introduced the longer KS4, just over half of schools responding to the survey made the change for all subjects at the same time (56%; Table 24); a third indicated that it had been a staggered process, taking place over a number of years, (35%). Of the schools that reported bringing in a three-year KS4 gradually, the most commonly

mentioned approach in the survey was to introduce it for science, maths, and English before rolling out to other subjects, although there were a wide variety of ways of implementing the change. There was also diversity in the number of subjects changed at a time with some starting with one or two subjects only and others rolling it out for most subjects from the beginning.

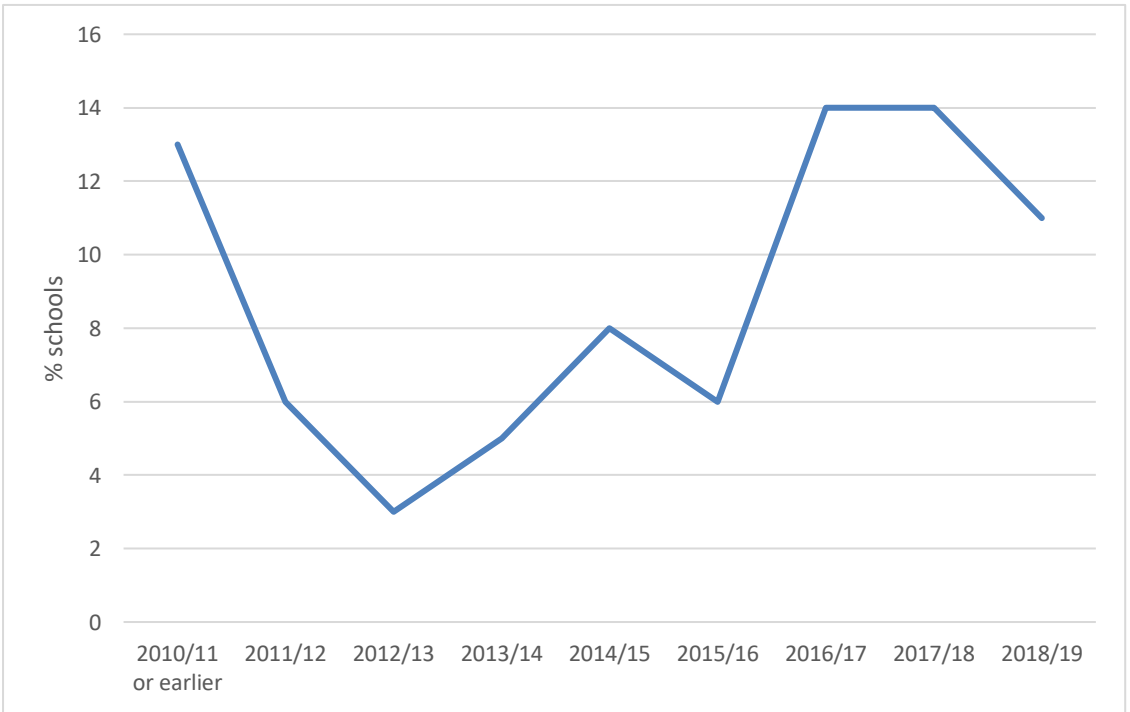
Table 24: Whether the change to a three-year KS4 was made for all subjects at the same time

Did all subjects switch to a three-year KS4 at the same time?	%
Yes	56
No	35
Not sure	5
No response	4

N = 263 schools.
A single response item.
Responses are limited to schools that begin KS4 in Year 9 for at least some pupils or subjects.
Source: NFER survey of publicly funded mainstream secondary schools, 2020.

Schools that had moved to a three-year KS4 tended to either be early adopters (introducing it in 2010/11 or earlier) or adopted a three-year KS4 from 2016/17 onwards - just after the new GCSE curriculum was introduced (in 2015) (Figure 26).

Figure 26: When did schools first start teaching KS4 from Year 9?



Overall N = 208. Chart and N excludes years with low n for disclosure reasons, those that said 'not sure', and non-responses.
Source: NFER survey of publicly funded mainstream secondary schools, 2020.

Over three quarters of schools with a two-year KS4 said they had always had a two-year model (78%; Table 25). A fifth (20%) indicated that they had had a different model at some point in the past. Almost all of these conversions back to a two-year KS4 model had taken place since 2016/2017 and the data suggested that this was becoming more common in the most recent years.

We also explored some of these issues in greater depth in the interviews. Where interviewees were based in schools that had changed the length of their KS3 or KS4 for one or more subjects, they were asked how the decision had been made and how this had been implemented. The findings are discussed below.

Table 25: Whether schools have had a two-year KS4 since the 2010/2011 academic year

Has the school had a two-year KS4 since 2010/2011?	%
Yes	78
No	20
Not sure	2

N = 142 schools.
A single response item.
Responses are limited to schools that begin KS4 in Year 10.
Source: NFER survey of publicly funded mainstream secondary schools, 2020.

Who was involved in the decision-making process?

In most cases, interviewees reported that it was the headteacher (or a previous headteacher) who had instigated the process of making a curriculum change. In most schools it appeared to be the senior leadership team (SLT), or individual designated members, who were responsible for implementing curriculum changes. For example, in one school a member of the school's SLT was tasked with reviewing the school's curriculum offer and for suggesting any changes.

A range of stakeholders were reported to have been consulted prior to any reforms being introduced. This usually included subject leads and classroom teachers and sometimes, but not always, parents and pupils. Two interviewees reported that governors were consulted, and in at least two academy schools, the multi-academy trust (MAT) was also involved. The range and depth of these consultations appeared to vary depending on the scale of the changes being proposed.

Where parents were consulted, their response was typically reported to be positive, but parents could also provide an element of challenge. For example, in one school, which planned to start teaching KS4 science in Year 9, the senior leader reported that some parents wanted the school to go even further and introduce all KS4 subjects in Year 9.

Interviewees reported that the lead-in time for implementing any curriculum changes ranged from approximately six to 18 months, depending on the scale of the changes being introduced.

What informed the decision?

Senior leaders gave examples of a range of activities to inform their decision-making; each of these were mentioned by only a few interviewees. These included (in no particular order):

- reviewing the latest research evidence;
- reviewing cross-curricular links;
- reviewing schemes of work;
- appointing a new member of staff to the SLT to create additional capacity to lead the curriculum review process;
- bringing in external consultants or specialists to audit existing provision; and
- visiting other schools to explore different implementation models.

How did schools ensure the new KS3 and KS4 model was implemented effectively?

Senior leaders reported drawing on a variety of evidence to assess and monitor the effectiveness of their curriculum reforms. This typically included:

- pupil performance data;
- formative assessment;
- lesson observations; and
- pupil and teacher feedback.

Some interviewees reported that prior to making any reforms they had mapped out what 'success' would look like and used data from the sources listed above to assess whether this had been achieved.

Curriculum breadth and depth

The issue of the breadth of the curriculum (the range of subjects offered) *and* depth of the curriculum (the level of detail within the content and concepts) was a key issue mentioned by participants in both the survey and the interviews. The survey findings reported above showed that this was the main reason that senior leaders in schools with a two-year KS4 operated that model, but it was also a key factor in the motivations of three-year KS4 schools. We explored this issue in further detail in the interviews; the findings are reported in this section split by model type as there appears to be a clear distinction in the way the different types of schools perceive this issue.

Schools operating a two-year KS4 model

Overall, schools operating a two-year KS4 curriculum model felt that their model maintained high levels of breadth and depth across both KS3 and KS4. A large proportion of these schools highlighted the importance of having a three-year KS3 for giving pupils the breadth and depth of knowledge, skills, and experience needed for success at KS4 and in later life. This is exemplified by the selection of quotations below:

> *'With the longer KS3 it is an opportunity to really embed some of the foundations of learning that they will need at KS4 but even in the longer term as well. A lot of these subjects they will never formally study again … this is a shame in terms of their broader development' (two-year KS4, interviewee).*

> *'The students require a breadth of subjects within Key Stage 3 for as long as possible to understand the wider world. Some have rarely left their estates in the last few years and have narrow horizons. We need to teach them as much as possible in Key Stage 3' (two-year KS4, interviewee).*

Furthermore, some schools with this model felt that giving pupils breadth and depth at KS3 was fundamental for achieving breadth and depth at KS4 as pupils narrow their options based on their KS3 experiences. For example, one senior leader reported that having three years in KS3 allows pupils to have a more varied, enriched experience of modern foreign languages and that, as a result, more of their pupils are choosing a language at KS4.

However, some senior leaders in schools operating a two-year model challenged the notion that they operated a 'two-year KS4 model' because they viewed the curriculum as a five-year continuum or as a 'spiral curriculum':

> *'Most areas of the curriculum are continuous, so Key Stage 3 and Key Stage 4 aren't so distinct [in] a five-year model. It's more that they stop learning some subjects … you lose out on that depth of knowledge. Ideally you'd study all subjects until end of Key Stage 4, but that's not possible or practical' (two-year KS4, interviewee).*

A small number of schools operating a two-year KS4 model acknowledged that in their school the curriculum model did reduce the number of options pupils could take in KS4 (compared to what the school might be able to offer in a three-year model). This group of senior leaders reported the increased depth and knowledge-rich content of the new GCSE qualifications meant that pupils were not able to choose as many subjects due to the time demands of each subject that needed to be met within two years:

> *'We actually made a conscious decision to reduce the number of options to allow more learning time in those subjects' (two-year KS4, interviewee).*

Schools with a two-year KS4 had differing opinions on how best to structure their options in order to give pupils breadth at KS4. Some felt that the EBacc gave pupils a broad, challenging curriculum and so structured their options process in order to maximize EBacc participation among pupils. However, some felt strongly that the EBacc was in fact restrictive for pupils and took away opportunities for pupils to study the arts and vocational courses in particular. For example, one senior leader reported that subjects like performing arts, PE, business studies, and technology subjects 'are well down

on numbers for choices because those taking all of the EBacc subjects can't take them'. Similarly, another senior leader commented:

> *'We make no reference to the EBacc when we discuss the options because we think this narrows the curriculum [particularly at the expense of the creative subjects]' (two-year KS4, interviewee).*

Schools operating a three-year KS4 model for all or some subjects

There was consensus among many of the schools operating a three-year KS4 model for all or some subjects that their curriculum model enabled greater breadth and depth at KS4. These senior leaders felt that this curriculum model created the time and space to maintain or increase the breadth of the curriculum at KS4 while also achieving depth of study. The senior leaders agreed that operating an extended KS4 model meant pupils had more time to study the KS4 content to greater depth while also being able to offer pupils more options subjects over three years. The headteacher in one school said:

> *'I feel we have quite broad curriculum at KS3, it's just that we end our KS3 in Year 8' (three-year KS4, all pupils, all subjects, interviewee).*

The evidence from the interviews suggests that schools with an extended KS4 were more likely to be operating innovative curriculum models. For example, a few schools structured their KS4 delivery to so that core subjects (English, maths, and science) were delivered over three years, while the options subjects were delivered intensively in a single year. As a result of the varied curriculum models these schools operated, there was a wide variety of approaches in how these schools structured their options process to accommodate breadth and depth at KS4. Some three-year schools advocated EBacc subjects but had additional options slots for creative and vocational subjects while others offered pupils an entirely free choice. The qualitative evidence suggests that because schools with an extended KS4 tended to offer pupils more options slots, these schools did not experience the same reduced participation in arts or vocational subjects as reported by schools with a two-year KS4 model.

Although schools operating a longer KS4 used the extra time in Year 9 in a variety of ways, schools had two main priorities for how this additional KS4 teaching time was utilised: enabling teachers to deliver the curriculum in greater depth or protecting time for revision in Year 11. For example, one school operating a three-year KS4 for English and maths delivered the KS4 content for these subjects in Year 9 and Year 10, using Year 11 solely for the purpose of revision and exam preparation. At a different school, pupils studied their KS4 content in greater depth and the courses embedded some KS5 preparation work and content into the KS4 curriculum. Finally, as discussed previously, some schools used the extra time to offer pupils an additional subject in order to offer pupils a broader, more varied KS4 curriculum.

Only a small minority of schools commented that by extending KS4 into Year 9 for some or all subjects, KS3 was restricted in terms of the breadth and depth it was able to achieve. Staff in other schools with a three-year KS4 reported that they had been creative in structuring their curriculum model to overcome the inevitable squeeze on curriculum time at KS3. For example, one school operated a model whereby pupils select four options in Year 8 and then select their final three options to pursue in KS4 at the end of Year 9.

Plans for the future

During the survey, we asked schools to indicate their future plans for KS3 and KS4, and specifically if they intended to change their model.

None of the 123 two-year KS4 schools responding to the second part of the survey said they planned to adopt a three-year KS4 model.[39] Reasons for retaining a two-year KS4 were predominately the same as those given above:

- to maintain or maximise curriculum breadth and depth at KS3 (35 schools; 28%);
- school has taken alternative steps, for example, curriculum review, five-year curriculum (15 schools; 12%);
- the importance of having a three-year KS3 to prepare for KS4 (11 schools; 9%);
- views that their current model works for their school (11 schools; 9%);

[39] As explained in the Methods section, the second part of the survey was optional.

- a three-year KS4 does not align with school ethos (ten schools; 8%);
- pupil readiness (ten schools; 8%).

In contrast to the two-year KS4 schools (that all planned to retain their current model), almost a fifth of the 198 schools with a three-year KS4 (that completed the second part of the survey) reported firm plans to move back to a two-year KS4 (19%; Table 26). An additional 16% had considered making the change but had not yet decided what to do. More than half (57%) planned to retain a three-year KS4.

Table 26: Do three-year KS4 schools have plans to start teaching KS4 from Year 10 in the future?

Are there plans to start teaching KS4 from Year 10 in the future?	**%**
Yes, this is planned	19
Considered, no firm plans	16
Considered, decided against it	26
Not considered	31
Not sure	5
No response	4

N = 198.
A single response item.
Responses are limited to schools that begin KS4 in Year 9 for at least some pupils or subjects and to respondents that completed part two of the survey.
Source: NFER survey of publicly funded mainstream secondary schools, 2020.

The frequently cited reason for changing back to a two-year KS4 was in response to the Ofsted inspection framework (mentioned by 15 schools planning to change back); this was also the most common factor mentioned by schools that were considering the change but had not made any firm plans yet (mentioned by 17 schools). Some of these schools *only* cited Ofsted as the driver for the planned change: in some cases it was directly related to an inspection, in others it was in response to the Ofsted Framework and in anticipation of being inspected.

This was a common theme in the interviews; several interviewees felt that the education inspection framework (Ofsted, 2019) pointed towards favouring a two-year KS4 because their understanding was that the quality of education judgement starts from Ofsted's premise that schools should offer their children a broad, rich curriculum. Several survey respondents described andecdotally that they had been told that it would not be possible to achieve a high rating from Ofsted with a three-year KS4 – this included information coming from other schools, independent inspection advisors (e.g. during mock inspections) and – less frequently - during inspections themselves. Respondents described a fear of being downgraded at their next inspection, or that they simply did not want to take the risk.

As noted in the introduction to this report, Ofsted has stated that it does not have a preferred curriculum model and that inspections expect schools to offer pupils an 'ambitious' curriculum with the opportunity to study a broad range of subjects across their whole time in secondary education (Harford, 2020; Ofsted, 2020a). However there appears to be a disconnect between this position, and the perception in at least some of the schools that participated in this research.

However, interviewees from a handful of three-year KS4 schools said that they had demonstrated the benefits of their model to Ofsted. Staff from schools that had achieved a 'good' or 'outstanding' rating with a three-year KS4 said that they had provided additional evidence demonstrating that they had maintained good breath within the curriculum, for example by varying the length of KS3 and KS4 by core or optional subject, or by ensuring a 'vibrant' extra-curricular offer with high pupil engagement. One interviewee from a school rated Outstanding while operating a three-year model explained:

> 'We keep reviewing it, and feel it does give our pupils the best experience…they [Ofsted] came in very sceptical about the model as a whole but they were impressed by what they saw and by what the pupils said about it, and the opportunities it gave [the pupils].' (three-year KS4 school, considered changing but decided against it, interviewee).

These three-year schools (whether they had had a positive experience with Ofsted or not) all emphasised that their curriculum had been designed with their pupil cohort in mind and remained convinced of the benefits of their approach for their pupils. One of the senior leaders interviewed explained their perception:

> '[There can be a] slightly reductionist view of KS3/4 … assuming that schools that choose to move some aspects of KS4 are doing it because they're trying to game league tables and things like that and I think there's been a failure to understand there are valid reasons [for this choice]' (three-year KS4, all pupils for some subjects, interviewee).

Other reasons given by schools that had committed to, or considered, changing back to a two-year KS4 (N = 37) included maximising the curriculum breadth and depth at KS3 (eight schools), the time needed to deliver the curriculum content (seven schools), and pupil readiness (six schools) in line with the rationale given by the two-year schools for their model. Some mentioned that they had conducted curriculum reviews in which they had identified that they did not need three years to teach the KS4 curriculum, which in turn facilitated the change. However, others described sacrifices they had made within the curriculum (for example, reducing the number of subjects studied at KS4) in order to maintain the minimum amount of time they felt core subjects required at KS4.

Three-year KS4 schools planning to retain their current model (N = 112) cited reasons already described above - in particular: meeting the needs of pupils (17 schools), the time needed to deliver the content (15 schools), and curriculum breadth/depth at KS4 (13 schools). The most commonly-cited reason for retaining a three-year KS4 was that their model was beneficial to their school in terms of pupil engagement, curriculum delivery, and outcomes (28 schools). These schools were also slightly more likely to mention that feedback from key stakeholders in their school supported their current model (12 schools planning to retain a three-year model):

> *'After a survey of staff, students, and parents, the overwhelming consensus was that a three-year KS4 was in the best interests of our students' (three-year KS4 school: retaining model, considered changing but decided against it, survey respondent).*
>
> *'These curriculum decisions were taken following a detailed consultation with our students and our parents and have had a positive impact.' (three-year KS4 school: retaining model, not considered changing to two-year KS4, survey respondent).*

By the time of the interviews, around 15 months after the survey took place, six of the schools interviewed had made a change to their curriculum - in all cases from a three-year to a two-year KS4. In almost all cases this change had been planned at the time of the survey and related to the issues described above (curriculum review, curriculum breadth and depth). The impact of Covid-19 and the disruption to education had not been the sole factor in any school's decision, although some schools mentioned that it had hastened the change. The schools felt the priority was to ensure that the students had a good grounding in KS3 content so that they have the foundations upon which to build in KS4. The disruption had, in their view, highlighted the importance of building students' 'foundational knowledge' and strengthened their conviction that their planned move to a two-year KS4 was the right step:

> *'[The students] have been in a situation where, over the last year or so, we have opened and closed, opened and closed. We wanted to give students more time to get to grips with their courses [before commencing study at GCSE]. It gives the students the opportunity to build a broad and balanced curriculum as they move into the later year groups' (two-year KS4, interviewee).*

Whether there was a specific influence of Covid-19 on the curriculum and what form that took is discussed in the next section.

Influence of Covid-19 on the curriculum

In the majority of schools, Covid-19 appeared to have entrenched their position on the best curriculum model for their pupils. Schools of both types felt that their current models would support their pupils to recover from the impacts of Covid-19 - this aligned with the relative strengths each group saw in their model. For example, some operating a three-year KS4 felt their model gave them flexibility to adapt their content delivery to meet their pupils' needs as required because they have more time to deliver the KS4 content. One senior leader commented: 'If anything, it's hardened our view that we need longer to prepare the pupils for their qualifications' (three-year KS4, all subjects, all pupils).

In contrast, some two-year KS4 schools felt that their three-year KS3 meant that pupils had the time and breadth of study to build skills and knowledge needed for KS4 as well as the subject experiences needed to select the options best suited to them for KS4.

> *'I'm fully in favour of a three-year KS3 for lots of reasons but probably it's been polarised I think with what's going on with Covid … Schools that are rushing through two years of KS3 with the context of having kids missing some school - I think that could be potentially a bit worrying for how they're going to approach GCSE and the knowledge gaps they may well have' (two-year KS4, interviewee).*

While the majority of schools acknowledged the impact of Covid-19 on pupils and the challenges it has presented for delivering their planned schemes of work, they largely did not view changing their curriculum model as the solution:

> *'The pandemic has had no real influence on my thinking. However, if anything, it would strengthen my view on the value of promoting creative subjects to pupils' (two-year KS4, interviewee).*

> *'While we have definitely seen the impact of Covid, understandably, we have certainly this academic year managed to continue delivering the curriculum to the planned programme throughout' (three-year KS4, all subjects, all pupils, interviewee).*

Only a few schools were planning to revise their model in response to the impacts of Covid-19. In one example, a school that operated a three-year KS4 model prior to the pandemic moved back to a two-year model as a result of the disruption. The lack of face-to-face teaching and poor engagement among some pupils were key considerations that contributed towards the school deciding to alter their curriculum model. As the senior leader explained:

> *'[It] is a change we've implemented due to the pandemic and the perceived outlook of Ofsted on the two-year KS3 versus the three-year KS3. Morally we could not say the children had covered enough content by the end of Year 8 because of the gaps … there's been ongoing closures and ongoing moves to online and remote learning' (two-year KS4, previously three-year KS4, interviewee).*

Staff in other schools felt the pandemic had given additional impetus to their plans or decision to change their curriculum model. For example, in one school that had made the decision to move from a three-year to a two-year KS4 model, the senior leader reported that: 'Covid contributed 30 to 40% to the decision, but the main reason for the change is that we want students to have grasped key concepts before commencing the KS4.' The staff in the school felt that pupils need more time in KS3 to develop their numeracy and literacy foundations, to benefit from a knowledge-rich curriculum, and to have the time to engage more with subjects like humanities and modern foreign languages. The senior leader explained that the pandemic had reinforced the need for this change because the disruption had highlighted the gaps in pupils' foundational knowledge that would usually be addressed during KS3.

Some schools were modifying other aspects of their curriculum delivery in response to the pandemic's impact, including the structure of the school day (such as moving to six 50-minute lessons a day), the time allocation given to each subject, or adapting their curriculum content.[40] For example, at one school, the senior leader explained that due to the amount of teaching time missed due to Covid-19 disruption, the timetable for the next Year 11 cohort would dedicate 60% of their timetable to core subjects:

> *'Every day they will do English, maths, and science … but it's certainly not ideal because it's not an engaging curriculum' (two-year KS4, interviewee).*

Some schools used the partial school closure periods to review their curriculum to ensure that the core knowledge and skills pupils are learning is explicit and ensure no time is lost or wasted, as the following senior leaders explained:

> *'We've really driven teams to go in at far greater depth than I've seen elsewhere about what you're learning, why you're learning, and what you need to be able to do on the back of it' (two-year KS4, interviewee).*

> *'Part of their review was to look at the key elements of their curriculum. You would hope a child who wasn't continuing with that subject to KS4, that they would know about' (two-year KS4, interviewee).*

[40] This reflected the findings of Nelson et al. (2021) who classified four approaches to the curriculum that schools were taking in response to the disruption caused by Covid-19.

Perceived impacts

The survey findings reported above showed that the main reasons for adopting a particular curriculum model were often related to how they would benefit the pupils. The interviews explored this in greater depth and the following sections on impacts and challenges on pupils and teachers are all based on the interviews with senior leaders.

Perceived impact on pupils

Interviewees identified a range of benefits for their pupils, as well as some challenges, relating to their schools' curriculum structures or implementation choices. Many of these perceived benefits (or challenges) reflect the reasons given by schools for their curriculum model.

Benefits

In most schools, progressing from KS3 to KS4 involved a degree of specialisation, with pupils often having to choose a smaller range of subjects for GCSE study (or equivalent). The perceived impact of this on pupils was a point of difference between several of the senior leaders, depending on whether or not their schools provided a predominantly two- or three-year KS4.

Those in schools with a two-year KS4 gave the following benefits for pupils:

- in their view the main benefit of increased time at KS3 was that pupils could study a broader range of subjects for longer, including the arts and creative subjects, which helped them to build their 'cultural capital'. One senior leader explained that pupils were able to develop 'a broader and richer understanding of the subjects they [eventually] drop, ensuring they also have the awareness they need as adults'. This, they argued, was especially important for disadvantaged students, a point that other interviewees agreed with. Another senior leader, who was also based in a school with a two-year KS4, argued that:

'*KS4 should never be a polishing exercise whereby you finish the content in Year 10 and then simply revise for an exam. How awful and what message is this giving children about the point of learning!*' *(two-year KS4, interviewee)*.

- having more time to develop the foundational subject knowledge and skills, which pupils would need to build upon at KS4:

'A lot of the focus goes on high stakes results but people forget that in order to get there you need a strong foundation in the subject and grasp [of] key content, and is one of the things that often gets neglected in schools' (two-year KS4, interviewee).

- pupils having more experience and maturity in Year 9, enabling them to make more informed options choices for KS4 study.

By contrast, those in schools which provided a predominantly three-year KS4 reported the following benefits:

- the main benefit for pupils was providing more time to cover KS4 curriculum content, undertake revision, and develop relevant skills/understanding. Several interviewees reported that this was particularly important for GCSE science (combined and triple), given the large amount of learning content. One senior leader suggested that the extra time at KS4 allowed pupils to develop a 'rhythm for learning' (three-year KS4 for some subjects), which allowed them to be more focused and productive in Years 10 and 11.
- it also allowed pupils to make changes to their options choices, including giving pupils more time to decide whether studying triple science was appropriate for them, while leaving open the option of changing to combined science.
- it also expanded access to KS4 fieldtrips and extra-curricular activities.

While all of the senior leaders we spoke to felt that their pupils' wellbeing needs were being met, interviewees' comments suggested this was being achieved in different ways, depending on the KS3/4 model adopted. For example, giving pupils more time at KS3 was reported to help build pupils' confidence and made them better prepared when it came to making options choices. By contrast, where GCSE exams were taken early (there were indications this was more likely

with a three-year KS4), it was suggested that this alleviated some of the stress in Year 11 as pupils were studying for fewer exams at this point.

There was little evidence to suggest that schools' different curriculum arrangements impacted differently on a pupil's workload. Rather, school decisions regarding the frequency of homework and assessments, which appeared to be largely independent of the length of KS3 or KS4, were more frequently reported to have an impact in this area. For example, in one school (two-year KS4), the number of assessment points per year had been reduced from six to two in an effort to reduce pupils' workloads. Whereas another senior leader (also two-year KS4) reported that they had encouraged all subjects to set homework. This was to create greater parity between subjects and to 'encourage pupils to appreciate the subjects' worth', but it also added to pupils' workloads.

Interestingly, most senior leaders did not think that the KS3 or KS4 model adopted led to any specific benefits (or challenges) for their disadvantaged pupils, distinct from other pupils in their school. Several leaders said that factors such as the provision of high quality, engaging teaching, dedicated time with teaching assistants or learning support assistants, and support with KS4 options choices were more important than the length of KS3 or KS4. Another explained that it was important to look closely at what disadvantaged pupils were learning to see whether there was any additional support that could be put in place. This interviewee explained that their school had appointed someone to look specifically at student transition points, 'with a specific focus on disadvantaged pupils to make sure we've made the right decisions for them' (two-year KS4). A small number of senior leaders, whose schools had changed their curricula to commence KS4 study in Year 9, reported this had led to their disadvantaged pupils making better progress, as one explained: 'we found our Progress 8 for disadvantaged children has significantly increased and is now more in line with what our non-disadvantaged students are achieving' (three-year KS4).

Another senior leader (three-year KS4, all subjects, all pupils) reported that improvements in attainment were possible as result of creating more time to support disadvantaged pupils with their KS4 programme of study. In contrast, a third senior leader, whose school provided a two-year KS4, countered this view. They argued that while pupils studying for three years at GCSE might get better grades, 'we feel [that] for our disadvantaged students the most important thing is that they are capable and competent learners to take their place in society at the end of their education and are motivated to do so' and that giving pupils a three-year KS3 was the best way to do this.

Other factors suggested by one or more senior leaders that were said to contribute to an engaging curriculum experience for pupils (and not just those who were disadvantaged) included:

- well-sequenced content that challenged pupils at every stage of learning;
- the use of mixed attainment classes, as distinct to setting or streaming; and
- limiting the amount of shared teaching so that pupils could more quickly develop positive working relationships with their teachers.

Challenges

Senior leaders identified relatively few challenges for pupils resulting from their KS3 or KS4 models. Where concerns were expressed, these were most frequently from the senior leaders of schools who provided a predominantly three-year KS4 and related to the lack of maturity of some Year 8 pupils when it came to making KS4 options choices. Some leaders reported that pupils were less able to choose less wisely at this age and that there was a danger that this could result in pupils struggling with some of their KS4 subject choices. However, not all interviewees agreed, with some arguing that poor choices could be made by pupils whether those choices were made in Year 8 or in Year 9. Others argued that low prior attainment posed a greater challenge to whether or not pupils would be able to access KS4 content successfully.

One senior leader, who worked in a school that provided a predominantly three-year KS4, felt that there was a danger that pupils could view their school as an 'examinations factory'. Another, who also worked in a school that provided a predominantly three-year KS4, reported that their Year 9 pupils could sometimes be overwhelmed by the GCSE curriculum. Adaptations to content, and careful sequencing, were reported to help mitigate against this.

Other challenges, which did not appear to be directly associated with a school's curriculum structures or implementation choices, and reported by one or more interviewees, included:

- academically able pupils, who were often good at a range of subjects, sometimes struggled to make KS4 options choices;
- parents sometimes encouraged their sons/daughters to choose subjects that staff felt were not appropriate for them (for example, triple science); and

- different subjects completed the delivery of course content at different times of the year, leaving some teachers with less time to work on revision and exam skills.

Perceived impact on teachers

Senior leaders had mixed views on the impacts of their KS3 and KS4 curriculum models on their teachers. These ranged from those who had no strong views to those who identified a range of benefits for their teachers - as well as some challenges.

Benefits

One or more senior leaders who worked in schools that provided a predominantly two-year KS4 argued that the additional time given to KS3 meant:

- teachers had more opportunity to think about curriculum design and the construction of learning at KS3, and to introduce content that was of specific interest to them or that they thought was relevant to their cohort; by contrast, the GCSE syllabus was reported to be more prescriptive;
- pupils were better prepared for KS4 study, which in turn helped facilitate better teaching and learning at this key stage; and
- pupils made more informed options choices in Year 9, which again helped facilitate better teaching and learning at KS4.

One or more senior leaders who worked in schools that provided a predominantly three-year KS4 argued that the additional time given to KS4 meant:

- teachers were no longer under pressure to deliver a challenging KS4 programme of study in two years;
- there were more opportunities for practical subjects, such as design and technology, to schedule in blocks of lessons; and
- teachers had more time to get to know their pupils at KS4 and to develop positive working relationships.

Interestingly, where schools had undertaken recent reviews of their curricula, the process - and not just the resulting changes - was, in some cases, reported to have led to benefits for teaching staff, such as 'reigniting staff passions' and encouraging staff to reflect on and evaluate current teaching practices and processes.

Challenges

Interviewees identified relatively few challenges for teachers resulting from their KS3 or KS4 models. Instead, they were more likely to identify common external challenges, such as the increased workload for their staff that had resulted from the introduction of reformed GCSEs in 2015 and, more recently, from the effects of Covid-19. Other reported challenges appeared to stem from schools' implementation choices or other reported features of secondary education, including:

- 'free' options choices leading to variability in pupil numbers for subjects in options blocks, which could lead to staffing challenges;
- low numbers of pupils opting to study a second language;
- pupil engagement sometimes dropping in Year 9 ('the Year 9 dip');
- accommodating teachers' requests for flexible working patterns and part-time hours; and
- non-EBacc subjects, such as performing arts, PE, and business studies, suffering from low pupil uptake at KS4.

Where concerns resulting from a particular KS3 or KS4 model were expressed, these tended to relate to the pressures of delivering a KS4 subject in two years (for those operating a two-year KS4 model). For those operating a three-year KS4, the main challenge was reported to be the increased workload associated with extended KS4 provision, including the need to adapt materials for use in Year 9. One senior leader suggested that with a three-year KS4 model there was also greater risk of teacher turnover causing issues for personalised learning; another reported that this model was more expensive to staff than a two-year KS4.

In addition, while the process of undertaking curricula reviews had led to some benefits for teachers (as reported in the section above), in many cases, it was also reported to have added to teachers' workloads (as they contributed to reviews and discussions), albeit for a short period of time.

Conclusion

Table 27: Key conclusions

Key conclusions
The research is unable to conclude that any differences in observed outcomes are due to the length of KS4. This is because it was not possible to achieve a strong match through the QED, with the two groups of schools on different GCSE performance trajectories prior to the KS4 length policy change. This caveat applies to the primary outcome (maths attainment), the secondary outcomes (English literature and 5 A*-C grades at GCSE), and to a subgroup analysis on the maths performance of everFSM pupils[41]. The independent evaluation team does not interpret any differences observed as causal - differences in outcomes may have been caused by other factors alternative to, or in addition to, changing the length of KS4.
The evaluator's recommendation is that schools should not make a decision about the length of their KS4 on the basis of the impact evaluation (QED) findings reported here.
Almost twice as many schools responding to the survey delivered KS4 over three years (for at least some subjects) rather than over two years.
Both shorter- and longer-length KS4 schools were offering - and pupils were taking - fewer qualifications at the end of KS4; the number of qualifications declined from a peak in 2011/2012 to the time of the survey in 2019/20. Analysis of curriculum breadth (a secondary outcome measure), and the IPE findings, indicated that this shift was driven by policy changes relating to school-level performance measures, such as the EBacc, and changes to the way that vocational qualifications contribute to school tables.Curriculum breadth and depth was a key consideration for schools irrespective of their curriculum model.
Schools running a three-year KS4 were motivated to do so by their views of the requirements of the new GCSEs and, to a lesser extent, to improve pupil engagement in Year 9. In contrast, schools that had maintained a two-year KS4 described their primary motivation as the importance of a strong curriculum and breadth of experience at KS3 when delivered over three years. Regardless of the length of their KS4, schools emphasised the importance of tailoring the curriculum to their specific intake and context to best support their pupils. Schools often considered KS3 and KS4 holistically and reviewed the sequencing and delivery of the curriculum across the five years, rather than as two distinct key stages.

Impact evaluation and IPE integration

Evidence to support the logic model

This project aimed to better understand the ways in which schools are organising and delivering KS4 to their pupils, and whether different models had a different impact on pupil attainment. Due to the challenges of response rates (influenced by conducting research during Covid-19 lockdowns) and a wider variety in the models operated than originally anticipated, the project shifted its focus to maths for a primary outcome, with English literature and achieving five A* to C GCSE grades (or equivalent) as secondary outcomes.

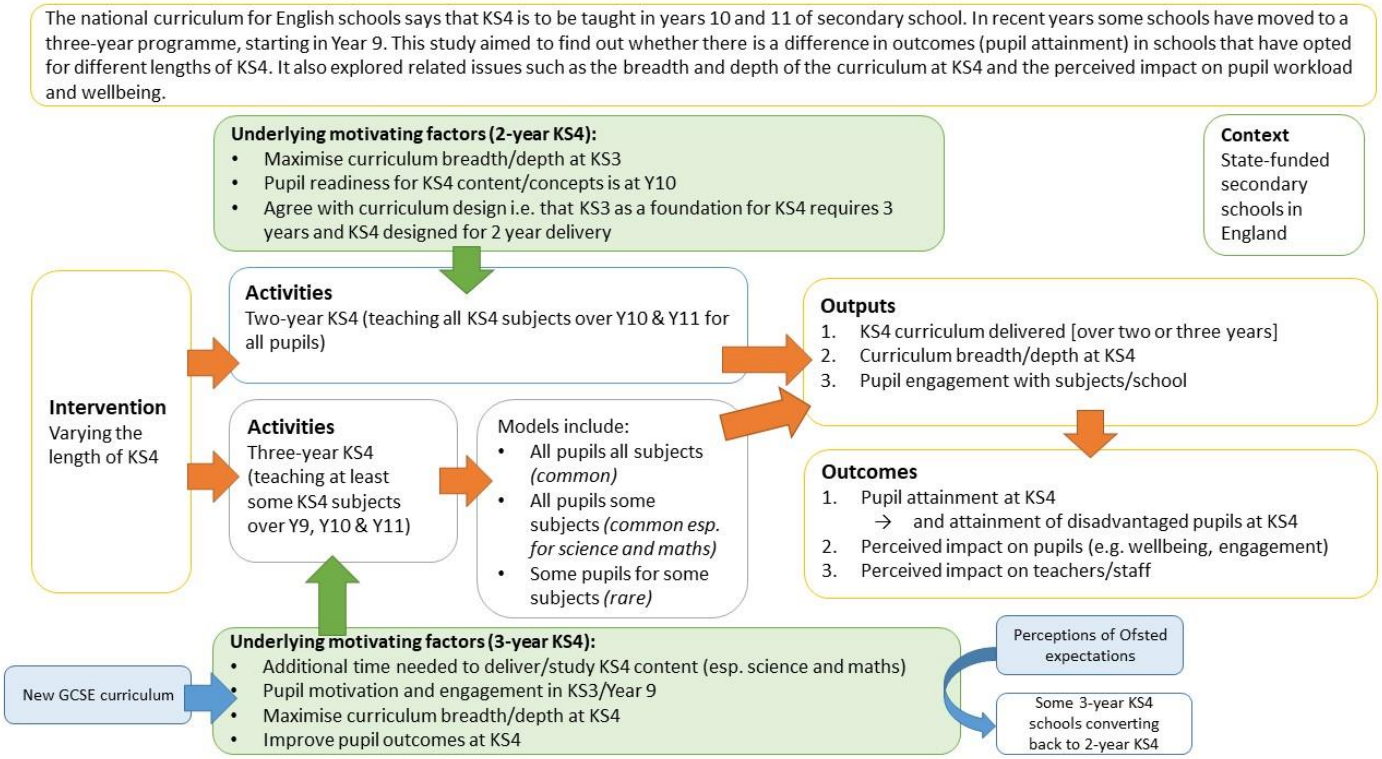
Almost two-thirds of the 405 schools that responded to the survey operated a three-year KS4 for at least some subjects. This is only slightly higher than that found by other similar research (for example, IFF Research, 2018; NFER 2019), however, as the survey sample was not representative of all schools, it is possible that schools that were more confident about their model responded to the survey. There was a wide variety in the way schools organised a longer KS4, varying from introducing some KS4 content in Year 9 for core subjects such as science and maths (which did not require an early options process) to a complete conversion from the start of Year 9, with options decided by pupils in Year 8. Because of this, the impact analysis focused on schools that reported a full additional year of KS4 maths, compared with those that reported KS4 starting in Year 10.

The project found some evidence, particularly from the IPE and curriculum breadth analysis, to support the activities, outputs, and outcomes of the original logic model and filled in some of the gaps around the moderators and motivators for different models. However although we report the findings of the impact analysis, we do not interpret the apparent effect on attainment to be causal due to the issues in the analysis, as explained below. An updated version of the logic model is shown in Figure 27. The mixed picture from the impact analysis suggests that the decision to change the length of KS4 is not one made in isolation as changes to performance were observed prior to the policy change. A similar observation was made in the curriculum breadth analysis, for example, that there was a divergence in the mean number

[41] 'EverFSM pupils' are those that have ever been eligible for free school meals (FSM).

of qualifications offered by three- and two-year schools that pre-dated the first cohort of pupils taking exams at the end of a three-year KS4. This was further supported by the findings from the interviews in which schools described the decision as part of a wider curriculum review, not solely on length of the curriculum. As the change to the length of the key stage is a significant change to make, which requires time and planning to implement, there are likely to be other changes (unobserved in the impact analysis but, for example, changes to teaching and learning) implemented in parallel or in preparation for the change in length of key stage. Such changes may be implemented more quickly and, therefore, potentially while earlier cohorts are still experiencing a two-year KS4. Furthermore, external policy changes (such as the change to GCSEs or the introduction of the EBacc) may also influence policy, strategy, or pedagogical changes in schools in other ways, and not solely as an influence on results via length of KS4.

Figure 27: Logic model updated based on the findings of the project



Interpretation

Overall, it is not possible to draw causal conclusions about the impact of different lengths of KS4 on attainment, based on the evidence reported here. This is discussed further in the next section: Limitations and Lessons Learned.

The impact analysis (n = 104) showed that, overall, pupils in schools with a two-year KS4 for maths had higher scores in both the pre-treatment and post-treatment periods. However, the interaction between treatment and time period (pre versus post policy switch) showed a small but significant effect suggesting that three-year schools had a larger improvement than two-year schools - but we do not conclude that this is a causal effect. This is because, importantly, the evidence suggested that the two-year and three-year groups being compared were not similar in terms of maths performance prior to the policy being introduced. The trajectories of the two- and three-year KS4 schools appeared to be converging anyway before the length of KS3 and KS4 was changed, and these trends may have continued with or without the policy change. Placebo tests were also run in the pre-treatment period, the results of which indicated non-parallel trends between the groups in this period. Taken together, this suggests that the parallel trends assumption is violated for the groups being analysed. This means we cannot conclude that any differences are due to the policy change and may instead be caused by other factors instead of, or in addition to, changing the length of KS4. A model that allows for linear differential trends between the groups is therefore likely to be more realistic: calculating this model produced a smaller and statistically non-significant interaction between treatment and time period. This exploratory, post-hoc analysis provides no evidence of a direct causal impact of the introduction of a three-year KS4 on GCSE maths scores.

The lack of confidence in the results of the primary outcome analysis is also relevant to the subgroup analysis (FSM pupils) and the secondary outcomes of English and five A* to C grades as these are comparing the same groups of matched schools as for maths. We summarise these results briefly here, but advocate caution over these findings. When looking at the impact on disadvantaged pupils, there was evidence to indicate that the maths performance of everFSM pupils in schools that converted to a three-year KS4 increased post treatment compared to everFSM pupils in schools

that kept a two-year maths KS4. Analysis by school-level disadvantage indicated a mixed and inconsistent picture. For English literature the three-year KS4 (maths) schools appear to improve post policy switch, but not as much as the pupils in two-year KS4 schools. When looking at performance more broadly, different lengths of KS4 were not associated with a pupil being more or less likely to achieve the five A* to C measure. We discuss the limitations to this analysis in further detail in the next section.

The survey and interview responses identified that the main reason for schools converting to a longer KS4 model was wanting additional time to study the content for the new GCSEs, in particular for maths and science.

The IPE found that in schools where a longer KS4 was operated for some - but not all - subjects, the most commonly selected subjects for a three-year KS4 were maths and science. Future research could explore this in further detail by matching on the performance of specific subjects and selecting different groupings for three- and two-year KS4 for English. Furthermore, if the dataset of KS4 models could be expanded to include many more schools, it would be possible to look at the five A* to C measure in schools that operate a three-year KS4 for all subjects, compared to two-years for all subjects. This is information that could be collected through the school census, and which could be added to the NPD for future analysis.

Motivating factors appeared to be clearly delineated between the two-groups (see revised logic model, Figure 27). The initial trigger for the shift to a longer KS4 in many cases appeared to be the new GCSE curriculum, which the three-year KS4 schools felt their pupils needed more time to be able to cover effectively. However the IPE analysis also indicated that lengthening the curriculum might only have been one of a number of measures that the schools may have implemented: the improvement was already happening before the change in length.

One of the key findings from the survey and interviews with senior leaders was the extent to which the three-year KS4 schools were considering changing back to a two-year KS4; indeed, several schools that were interviewed had already moved back to a two-year KS4 since the survey. The most frequently cited reason for this was the perception that Ofsted would look more favourably on their school during upcoming inspections, despite assurances from Ofsted that a three-year KS4 implemented well can also be ‘good’ or ‘outstanding’. In contrast, evidence from the survey showed that none of the two-year schools were considering lengthening their KS4 curriculum.

As the interviews were delayed due to the Covid-19 pandemic, we also took the opportunity to explore whether schools’ thinking about the length of KS4 has changed as a result of the disruption. In most cases, senior leaders continued to report that, in their view, their school’s model would support pupils well, but a minority of three-year schools were planning to move back to a two-year KS4 - and a corresponding three-year KS3 - to increase the time for pupils to study foundation content before moving into KS4.

The project also explored the breadth and depth of the curriculum. Overall, the mean number of qualifications entered by pupils had declined slightly over time, from almost 11 to 9.5 in 2018/2019. This appears to have been driven mainly by a reduction in non-EBacc subjects and was similar in two- and three-year schools. In terms of the breadth and variety of subjects offered by schools, the range has dropped considerably, again driven by a decline in the offer of non-EBacc qualifications. In contrast, qualifications eligible for the EBacc increased slightly. Any changes to the breadth or range of subjects and qualifications offered does not appear to be linked to a change in the length of KS4 as the decline is apparent in both school types, and in the case of three-year KS4 schools, the decline pre-dates the lengthening of KS4. The concern to maintain breadth and depth was reflected in the survey responses and interviews: school leaders from both types of school emphasised its importance, although whether the focus of breadth and depth should be in KS3 or in KS4 tended be reflected in the school’s length of KS4 (two-year and three-year KS4, respectively).

In addition to our comparisons of the curriculum breadth and depth by length of KS4, our initial exploration of overall changes to the types of qualifications entered at KS4 varied over time and produced some interesting findings, particularly in the context of the Wolf report (2011). The number of qualifications entered per school appears to have steadily declined since a peak in 2011/2012, with the bulk of the reduction largely in ‘non-EBacc’ qualifications. Entries for non-GCSE qualifications took a particular tumble.

Overall, this research found that schools had used their freedoms to organise the curriculum in many different ways, often quite unique to the school. The schools that took part in this research highlighted a number of key points for organising the curriculum effectively.

- As with individual pupils, schools are unique in terms of their intake, circumstances, and organisation. Schools do not perceive a one-size-fits-all approach to KS3 and KS4 delivery: participating schools had tailored their curriculum - to a lesser or greater extent - with the aim of best serving their pupils.

- Schools did not view a particular model as inherently ‘right’ or ‘wrong’; rather, they sought a ‘right approach for them and their pupils’.

- Several of the participants said they considered KS3 and KS4 as a whole and reviewed the sequencing and delivery of the curriculum across the five years, rather than as two distinct key stages.

- In deciding on their curriculum, schools consulted with stakeholders, reviewed performance, and considered other, non-academic aspects such as readiness of pupils.

- Curriculum breadth and depth can be achieved across KS3 and KS4 in two-year and three-year KS4 schools. School staff found it more challenging to evidence the benefits of this in a three-year KS4.

Limitations and lessons learned

As discussed above, we do not have high confidence in the results of the impact evaluation due to the apparent violation of the parallel trends assumption. There may be several reasons for the differences in pre-treatment performance trends, which may be incompatible with the assumptions made in this analysis. Some of these are discussed here.

- The assumptions and decisions made during the matching may not be quite right. First, the decision to impose an exact match by region and attainment was likely too restrictive given the small number of potential control schools and led to other variables not contributing to the match. Also there may have been a more suitable attainment measure available for the match. Schools in the two-year KS4 group were matched to a list of three-year KS4 schools using KS4 Attainment 8 progress scores rather than a maths measure. Attainment 8 was chosen for the match as originally we were due to have co-primary outcomes of maths and English. We felt that a broad attainment measure would be suitable given the inclusion of the secondary outcomes of English and five A* to C grades, however, we recognise that it would be good in future research to be able to explore the option of different matches by subject to try and improve the pre-treatment match for each analysis.

- The propensity score matching (PSM) was conducted using school-level variables (before the data was linked to the National Pupil Database) because we were working in the context of a challenging recruitment, and needed to understand whether we had sufficient schools for the analysis. The match was, therefore, conducted on school-level observable characteristics only, prior to obtaining access to the pupil-level data from the NPD. We recognise it would have been better if we could have run the match once we had access to the pupil-level data, on a combination of pupil- and school-level characteristics. Furthermore, as we were then restricted to only the data for the schools in the original match in NPD, it was not possible to re-run the matches at a later stage. Future research may benefit from running the PSM with pupil-level data from the NPD to include the pre-conversion performance of the outcome measure as we found some pre-conversion differences in performance between the two- and three-year schools in the maths performance model. This is likely related to the difference in the variable used for the PSM and the outcome measure.

- The analysis was limited by the number of schools that responded to the survey, which especially restricted the ability of the matching procedure to produce well-balanced groups (compounding limitations of the matching described above). Furthermore, there will be an element of self-selection of schools that completed the survey, although we subsequently selected schools for the analysis based on a number of eligibility criteria. This analysis was conducted on the available matched sample that was eligible for analysis, which was not representative of schools nationally. Therefore, we can see what happened in the case of different decisions about the KS4 curriculum in schools in this analysis but it is difficult to generalise to the school population more widely. Future research may wish to expand the dataset for analysis to address this gap. Although our main analysis only included schools that switched to a three-year model and did not switch back, compared to ‘never switchers’, there is also the additional question of what happens in schools that move back to a two-year KS4 after a period of operating KS4 over three years. The apparent flux in schools moving between models would need to be captured carefully to understand the drivers behind any impact.

- We had to make an assumption about when the decision would have been made in schools and apply this across the three-year group. We set this timepoint as the year before the first Year 9 cohort started their KS4 study, which we still feel is reasonable. The timing of a school’s decision to change by necessity means that schools would have been making changes in preparation for this change from around four years prior to the first Year 11 results. We note that the first cohort to study KS4 over three years would not be operating in a vacuum within the school; the last cohort to study under a two-year KS4 may also experience some effects related to the change - for example, different timetabling, possible changes to the syllabus and teaching methods, or resources in preparation for the three-year KS4: all of these things

may also filter out into that last 'two-year KS4' cohort, and perhaps even the cohort before. In other words, there is no clear discontinuity in the real world, even if it appears there may be one in the data.

- Furthermore, the policy change may have only been part of a suite of changes or policies implemented by the schools, evidenced by the apparent continual improvement pre-dating the curriculum change. Evidence from the 'free-response' questions in the survey and the interviews with headteachers and senior leaders in schools indicated that, as might be expected, decisions about the length of KS3 and KS4 were not made in isolation: other determining factors included the breadth and depth of the curriculum content and external drivers such as school accountability measures.

At this point, it may be helpful to reflect on whether it is plausible that matching can ensure a sample of two-year and three-year schools with parallel pre-trends given the policy question we are looking at and the data used. According to the evidence from the interviews and the open survey questions, past attainment patterns are one of the main things driving schools' decisions about the length of their key stages and, as described above, the decision was often part of other changes to policy and practice within the school which may also affect grades. Therefore, similar pre-treatment performance in two- and three-year schools during the years immediately before the change might not be a reasonable assumption. Adjusting the matching procedure used for this analysis in future research may yield more illuminating insights, but we also advocate considering whether a different type of QED design that does not rely on parallel trends may be helpful here.

The analysis was also limited by the number of years of attainment data available. Additional data may even out some of the extreme fluctuations in the data (for example, *c*-9 to *c*-8 in Figure 4a) with the addition of data for the analysed cohorts. Considerations in any future analysis would need to include the use of teacher assessed grades in place of the usual assessment arrangements in 2020 and 2021.

In terms of classifying schools into two- and three-year KS4 schools, the proforma survey appeared to work well. There were, however, a small number of schools that viewed themselves as adopting a particular model but in subsequent interviews it became apparent that the situation was not so clear cut. The most common example was schools with a two-year KS4 that started introducing KS4 content in the second half of the summer term of Year 9. This did not affect the impact analysis, as only schools with a full third year were included in the models, but it is a reminder of the nuances and adaptations that schools are making within the curriculum that might not be immediately apparent.

Rather than relying on additional primary data collection, by adding the length of KS3 and KS4 to the school census data collection so that it is regularly part of the NPD datasets would allow more comprehensive analysis and address several of the limitations outlined above.

Future research and publications

This report outlines the findings for a subset of secondary schools in England, and there are a number of ways that the research could be expanded in future studies, for example:

- increasing the number of schools in the analysis by determining which KS4 model they use;
- re-running the matching process with different parameters, in particular consider using pupil-level data for the match;
- re-running the matching process based on other variables of interest (for example, different subjects such as science) to correspond to the subject in the outcome measure;
- investigating the impact of different lengths of key stage on other subjects, in particular science, which was chosen by many of the schools that had opted for a longer KS4 for some subjects;
- where multiple cohorts are included in the analysis, further exploration into the differences by cohort;
- widening the outcome of interest to include, for example, wider cognitive outcomes and non-cognitive outcomes, engagement, and attitudes to learning; and
- exploring the impact of a three-year KS4 on breadth and depth of the curriculum at KS3.

Originally, we aimed to include an element of pupil voice through focus groups during visits to schools, however this was dropped due to the Covid-19 restrictions and burden on schools. Future research should explore ways to include the views of pupils and their form or pastoral staff.

References

Andrews, J., Perera, N., Eyles, A., Heller Sahlgren, G., Machin, S., Sandi, M. and Silva, O., 2017. *The impact of academies on educational outcomes*. [online] Available at: <https://epi.org.uk/wp-content/uploads/2017/07/EPI_-Impact_of_Academies_Consolidated_Report-.pdf> [Accessed 31 October 2022].

Austin, P.C., 2011. An Introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), pp.399–424. https://doi.org/10.1080/00273171.2011.568786.

Bates, D., Mächler, M., Bolker, B. and Walker, S., 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), pp.1–48. https://doi.org/10.18637/jss.v067.i01.

van Buuren, S. and Groothuis-Oudshoorn, K., 2011. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), pp.1–67. https://doi.org/10.18637/jss.v045.i03.

Department for Education, 2014. *The national curriculum in England: key stages 3 and 4 framework document*. [online] Available at: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/840002/Secondary_national_curriculum_corrected_PDF.pdf> [Accessed 31 October 2022].

Department for Education, 2021. *Teaching a broad and balanced curriculum for education recovery*. [online] p.36. Available at: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1033448/Teaching_a_broad_and_balanced_curriculum_for_education_recovery.pdf> [Accessed 31 October 2022].

Education Endowment Foundation, 2013. *Pre-testing in EEF evaluations*. [online] Available at: <https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/Pre-testing_paper.pdf> [Accessed 31 October 2022].

Education Endowment Foundation, 2015. *Intra-cluster correlation coefficients*. [online] Available at: <https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/ICC_2015.pdf> [Accessed 31 October 2022].

Gill, T., 2018. *The impact of the introduction of Progress 8 on the uptake of qualifications in English schools – an update for 2016/17*. [online] p.31. Available at: <https://www.cambridgeassessment.org.uk/Images/519667-the-impact-of-the-introduction-of-progress-8-on-the-uptake-of-qualifications-in-english-schools-an-update-for-2016-17.pdf> [Accessed 31 October 2022].

GL Assessment, 2019a. *Closing the gap: the link between low self-esteem and poor reading attainment*. [online] Available at: <https://www.gl-assessment.co.uk/news-hub/research-reports/closing-the-gap/> [Accessed 31 October 2022].

GL Assessment, 2019b. *Dangerous diet: how exam rations endanger a broad and balanced curriculum*. [online] Available at: <https://www.gl-assessment.co.uk/news-hub/research-reports/dangerous-diet-how-exam-rations-endanger-a-broad-and-balanced-curriculum/> [Accessed 31 October 2022].

Harford, S., 2020. Making curriculum decisions in the best interests of children. *GOV.UK*. Available at: <https://educationinspection.blog.gov.uk/2020/01/09/making-curriculum-decisions-in-the-best-interests-of-children/> [Accessed 1 November 2022].

Ho, D., Imai, K., King, G. and Stuart, E.A., 2011. MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8), pp.1–28. https://doi.org/10.18637/jss.v042.i08.

IFF Research, 2018. *The school snapshot survey: winter 2017*. [online] Available at: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/727807/The_school_snapshot_survey_winter_2017.pdf> [Accessed 31 October 2022].

Kleinman, K., Sakrejda, A., Moyer, J., Nugent, J., Reich, N. and Obeng, D., 2021. *clusterPower: Power Calculations for Cluster-Randomized and Cluster-Randomized Crossover Trials*. Available at: <https://CRAN.R-project.org/package=clusterPower> [Accessed 31 October 2022].

Little, T.D. ed., 2013. *The Oxford handbook of quantitative methods*. New York: Oxford University Press.

Middlehurst, T., 2020. *Should key stage 3 be restored to the full three years?* [online] SSAT. Available at: <https://www.ssatuk.co.uk/blog/should-key-stage-3-be-restored-to-the-full-three-years/> [Accessed 31 October 2022].

National Foundation for Educational Research, 2019. *Education inspection framework 2019 - NFER response*. [online] p.13. Available at: <https://www.nfer.ac.uk/media/3429/education_inspection_framework_2019_-nfer_response.pdf> [Accessed 31 October 2022].

Nelson, J., Lynch, S. and Sharp, C., 2021. *Recovery during a pandemic: the ongoing impacts of Covid-19 on schools serving deprived communities*. [online] p.35. Available at: <https://www.nfer.ac.uk/media/4614/recovery_during_a_pandemic_the_ongoing_impacts_of_covid_19_on_schools_serving_deprived_communities.pdf> [Accessed 31 October 2022].

Neumann, E., Gewirtz, S., Maguire, M. and Towers, E., 2020. Neoconservative education policy and the case of the English Baccalaureate. *Journal of Curriculum Studies*, 52(5), pp.702–719. https://doi.org/10.1080/00220272.2019.1708466.

Neumann, E., Towers, E., Gewirtz, S. and Maguire, M., 2016. *A curriculum for all? The effects of recent Key Stage 4 curriculum, assessment and accountability reforms on English secondary education*. [online] Available at: <http://downloads2.dodsmonitoring.com/downloads/Misc_Files/KingsCollege141116.pdf> [Accessed 31 October 2022].

Noden, P., Rutt, S., Schagen, S. and West, A., 2007. *Evaluation of the two year Key Stage 3 project*. [online] Available at: <http://eprints.lse.ac.uk/13767/> [Accessed 31 October 2022].

Ofsted, 2015. *Key Stage 3: The wasted years*. [online] Available at: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/459830/Key_Stage_3_the_wasted_years.pdf> [Accessed 31 October 2022].

Ofsted, 2018. *An investigation into how to assess the quality of education through curriculum intent, implementation and impact*. [online] Available at: <https://www.gov.uk/government/publications/curriculum-research-assessing-intent-implementation-and-impact> [Accessed 1 November 2022].

Ofsted, 2019. *Education inspection framework (EIF)*. [online] GOV.UK. Available at: <https://www.gov.uk/government/publications/education-inspection-framework> [Accessed 1 November 2022].

Ofsted, 2020. *The annual report of Her Majesty's Chief Inspector of education, children's services and skills 2019/20*. [online] Available at: <https://www.gov.uk/government/publications/ofsted-annual-report-201920-education-childrens-services-and-skills/the-annual-report-of-her-majestys-chief-inspector-of-education-childrens-services-and-skills-201920> [Accessed 1 November 2022].

Ofsted and Spielman, A., 2017. *HMCI's commentary: recent primary and secondary curriculum research*. [online] GOV.UK. Available at: <https://www.gov.uk/government/speeches/hmcis-commentary-october-2017> [Accessed 1 November 2022].

Ofsted and Spielman, A., 2018. *HMCI commentary: curriculum and the new education inspection framework*. [online] GOV.UK. Available at: <https://www.gov.uk/government/speeches/hmci-commentary-curriculum-and-the-new-education-inspection-framework> [Accessed 1 November 2022].

Parameshwaran, M. and Thomson, D.J., 2015. The impact of accountability reforms on the Key Stage 4 curriculum: How have changes to school and college performance tables affected pupil access to qualifications and subjects in secondary schools in England? *London Review of Education*, 13(2), pp.157–173. https://doi.org/10.18546/LRE.13.2.13.

Priestley, M. and Shapira, M., 2017. *Narrowing the curriculum? Contemporary trends in provision and attainment in the Scottish Curriculum*. [online] Available at: <https://dspace.stir.ac.uk/bitstream/1893/25879/1/ECER2017_Narrowing%20the%20Curriculum_final%20%28002%29.pdf> [Accessed 1 November 2022].

Rambachan, A., 2022. *HonestDiD*. Available at: <https://github.com/asheshrambachan/HonestDiD> [Accessed 1 November 2022].

Rambachan, A. and Roth, J., 2022. *A more credible approach to parallel trends*. [online] Available at: <https://jonathandroth.github.io/assets/files/HonestParallelTrends_Main.pdf> [Accessed 1 November 2022].

Roberts, J., 2019. Exclusive: Heads fear Ofsted will penalise three-year GCSEs. *TES magazine*. [online] 9 Jan. Available at: <https://www.tes.com/magazine/archive/exclusive-heads-fear-ofsted-will-penalise-three-year-gcses> [Accessed 1 November 2022].

Rutt, S., Poet, H. and Dirie, A., 2020. *School choices: What works at Key Stage 4 in terms of improving GCSE outcomes, two or three years of study?* [online] Available at:

<https://d2tic4wvo1iusb.cloudfront.net/documents/projects/SC_what_works_at_KS4_-_Study_Plan_NFER_Final_version_2.pdf> [Accessed 1 November 2022].

Stuart, E.A., 2010. Matching methods for causal inference: a review and a look forward. *Statistical Science*, 25(1), pp.1–21. https://doi.org/10.1214/09-STS313.

Teerenstra, S., Eldridge, S., Graff, M., De Hoop, E. and Borm, G.F., 2012. A simple sample size formula for analysis of covariance in cluster randomized trials. *Statistics in Medicine*, 31(20), pp.2169–2178. https://doi.org/10.1002/sim.5352.

The R Foundation, 2022. *The R Project for statistical computing*. [online] Available at: <https://www.r-project.org/> [Accessed 1 November 2022].

Wolf, A., 2011. *Review of vocational education – The Wolf report*. [online] Available at: <https://www.stem.org.uk/resources/elibrary/resource/31837/review-vocational-education-wolf-report#:~:text=Review%20of%20Vocational%20Education%20-%20the%20Wolf%20Report,to%20further%20education%20or%20prospects%20for%20meaningful%20employment.> [Accessed 1 November 2022].

Appendix A: Security classification of trial findings

OUTCOME: *GCSE Mathematics*

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5	Randomised design	<= 0.2	0-10%					
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%	4				
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%			Adjustment for threats to internal validity **Minus 2 padlocks**		
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					2
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Risk rating	Comments
Threat 1: Confounding	High	There is a high-level of imbalance in almost all observable characteristics between the groups of two- and three-year KS4 schools at baseline, even after matching. The authors find a lack of evidence of parallel trends prior to the change in KS4 length. Given this, it seems likely that there will be imbalance in unobservable characteristics. The authors make regression adjustments, but this requires strong assumptions on linearity.
Threat 2: Concurrent Interventions	High	The authors identify through the implementation and process evaluation that many schools made the change to a 3-year KS4 programmes as part of a wider suite of other policy changes. The authors are unable to account for this in the analysis.
Threat 3: Experimental effects	Low	There is no risk the control will have behaved differently as a result of participation, given the treatment was reported by schools after the fact.
Threat 4: Implementation fidelity	Moderate	Implementation fidelity is explored, however there is some uncertainty about the extent to which there is a sharp difference in approach between 3- and 2-year KS4 schools and the extent to which there was a sharp change in length of KS4 from one cohort within a school to the next.
Threat 5: Missing Data	Low	The amount of missing data is quite small as there is a reliance on administrative data and analyses that account for missing data are similar to complete-cases analyses.
Threat 6: Measurement of Outcomes	Low	The primary outcome is GCSE Mathematics scores, which are known to be reliable and have external validity.

Threat 7: Selective reporting	Low	While the trial was not registered as it is a quasi-experimental study and EEF did previously require registration of quasi-experimental study, the protocol with pre-specified analysis plan was published on the EEF website before data was accessed and analysed. The analysis closely follows the pre-specified analysis plan, with deviations well-justified.

- **Initial padlock score:** 4 – Difference-in-differences design with low attrition and MDES
- **Reason for adjustment for threats to validity**: 2 padlocks – This is due to a high level of risk of confounding, together with a high risk that 3-year KS4 schools made other changes alongside the main treatment of interest (concurrent interventions).
- **Final padlock score:** initial score adjusted for threats to validity = - 2 padlocks

Appendix B: Effect size estimation

Appendix table 2: Effect size estimation

Outcome	Parameter	Parameter estimate (adjusted)	SE of parameter estimate	Pooled variance s^2
GCSE mathematics scores	Difference-in-differences coefficient	0.076	0.012	2.68

Further appendices:

Please submit any further appendices as a separate document of technical notes. We will be publishing these as a separate document, to reduce the length of reports.

This document is available for download at https://educationendowmentfoundation.org.uk

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

https://educationendowmentfoundation.org.uk
@EducEndowFoundn
Facebook.com/EducEndowFoundn