



Education
Endowment
Foundation

English Mastery

Evaluation report

September 2026

Martin Culliney, Karen Daniels, Lewis Clark and
Marie Helks

**Sheffield
Hallam
University**
Knowledge Applied



The Education Endowment Foundation is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, colleges, and early years settings to improve teaching and learning for 2- to 19-year-olds through better use of evidence.

We do this by:

- **summarising evidence**—reviewing the best available evidence on teaching and learning and presenting this in an accessible way;
- **finding new evidence**—funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socioeconomically disadvantaged backgrounds; and
- **putting evidence to use**—supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set up in 2011 by the Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were reendowed with an additional £137m from government, allowing us to continue our work until at least 2032.

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



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



0207 802 1653



info@eefoundation.org.uk



www.educationendowmentfoundation.org.uk



Department
for Education



Table of contents

About the evaluator	3
Executive summary	4
Introduction	6
Methods	13
Impact evaluation	24
Implementation and process evaluation.....	35
Conclusion	54
References	57
Appendix A: EEF cost rating	59
Appendix B: Security classification of trial findings.....	60
Appendix C: Effect size estimation	62
Further Appendices	63

About the evaluator

This project was independently evaluated by the Sheffield Institute of Education, Sheffield Hallam University. As a member of the Education Endowment Foundation evaluator panel for more than ten years, the evaluators have completed numerous randomised controlled trials and other studies for the EEF and other organisations including government departments, charities, and other policymakers.

The project email address is EnglishMastery@shu.ac.uk

Corresponding author: Martin Culliney (M.Culliney@shu.ac.uk).

Acknowledgements

Nick Speed, Tom Kidston, and Linda Bray all provided valuable administrative support. Sean Demack advised on the evaluation design. Martin Illingworth conducted data collection and analysis for the fieldwork visits.

We would also like to thank the English Mastery development team at Ark Curriculum, the teachers and pupils who took part in the evaluation, and colleagues at the EEF who were involved in the trial.

Some of the analysis was undertaken in the Office for National Statistics Secure Research Service using data from the ONS and other owners. This does not imply the endorsement of the ONS or other data owners.

Executive summary

The project

English Mastery aims to improve attainment in English by providing a substantial and ongoing programme of curriculum-integrated professional development for teachers. Developed by Ark Curriculum Plus, it is delivered through in-class teaching during regular timetabled English lessons as a whole-curriculum programme with two pathways: the traditional curriculum for pupils working at the expected level for their age and the foundation curriculum for those below the age-expected level. The programme involves a cumulative approach to curriculum design including bespoke assessments. The three key elements in the curriculum are Literary Heritage, Writing Mastery, and Reading for Pleasure. Schools nominate a member of their English department to become the English Mastery Lead (EML) and lead the implementation of the programme in their school. The EML leads weekly co-planning sessions to support teachers to plan their implementation of the curriculum together. It was tested thro. For this trial, participating pupils were in Year 7 and Year 8 (aged 11 to 13) and programme delivery took place over two school years. Recommended delivery time was three hours of Literary Heritage per week, two hours of Writing Mastery one week, alternating with one hour the following week. Reading for Pleasure was recommended for one hour in alternate weeks.

A total of 72 schools were recruited to the trial; 61 completed baseline and endline assessments leading to 3,780 pupils in the analysis sample. The impact evaluation design was a two-arm, three-level efficacy trial with pupils clustered into classes and classes clustered into schools. The primary outcome was English attainment. The implementation and process evaluation (IPE) comprised fieldwork visits to ten schools in the first and second delivery year and a teacher survey at baseline, mid-intervention (at the end of Year 7), and post-intervention (at the end of Year 8). Schools were recruited in two cohorts: delivery ran from Autumn 2022 to Summer 2024 for Cohort 1 and from Autumn 2023 to Summer 2025 for Cohort 2. The trial received funding from the Department for Education's Accelerator Fund.

Table 1: Key conclusions

Key conclusions	
1.	Pupils in English Mastery schools made no additional months' progress in English, on average, compared to pupils in other schools. This result has a low to moderate security rating.
2.	The trial aimed to evaluate the impact of English Mastery among children eligible for Free School Meals (FSM). There was a smaller number of pupils in this group and these results may have lower security than the overall findings. Among children eligible for FSM, those in English Mastery schools made two fewer months' progress in English, on average, compared to those in other schools.
3.	Compliance with the intervention was low with only ten of the 30 intervention schools deemed fully compliant, meaning most pupils did not receive the programme as intended. There was a sense that the timetabling requirements for planning and training were demanding. Some schools reported that time constraints impeded successful implementation, including difficulties in finding time to deliver full curriculum coverage and to schedule co-planning sessions, assessments, and training.
4.	Teachers suggested that English Mastery had a positive impact on their subject knowledge, particularly for teaching technical accuracy and grammar, and on workload.
5.	Teachers and pupils reported that English Mastery had a positive impact on pupil English skills and pupils reported increased self-confidence in English skills. Teachers also reported that pupils benefited from the challenge of the Literary Heritage and Writing Mastery strands.

EEF security rating

These findings have a low to moderate security rating. This was an efficacy trial, which tested whether the intervention worked under developer-led conditions in a number of schools. Although the trial was a well-designed two-armed randomised control trial, it was not as well-powered as originally intended due to recruitment challenges. The trial was not powered to detect an effect of the magnitude found in the study, adding uncertainty to the results. Additionally, pupil-level attrition was significant, with 27.6% of pupils in the intervention group and 26.6% of pupils in the control group not included

in the final analysis. Low implementation fidelity and the possibility that other interventions or experimental effects were happening at the same time further decreased the security of the findings.

Additional findings

While this trial does not provide evidence that English Mastery improves attainment, the impact evaluation was not able to generate high-quality evidence. Pupils in English Mastery schools made, on average, no additional months' progress compared to those in the control group. The possible impact of this programme also includes small negative effects of two months less progress and positive effects of up to one month of additional progress, reflecting uncertainty about the headline result. This is compounded by the diverging results from the two delivery cohorts. The first cohort delivering English Mastery was associated with a negative effect on attainment, whereas for the second cohort no effect was found (with possible impacts including positive and negative effects). There is no clear evidence to explain the differences in results making it difficult to summarise programme impact on the whole. Among children eligible for FSM, those in English Mastery schools made two fewer months' progress in English, on average, compared to those in other schools. These results may have lower security than the overall findings because of the smaller number of pupils. However, compared to the overall sample, attrition was lower, and the impact ranged from zero to four months less progress. There was no evidence from the IPE to account for why FSM pupils made less progress. The evidence from the trial was further weakened by low compliance, with some schools reporting that time constraints impeded successful implementation. Full implementation involved a significant change in practice that was challenging for some schools. Specific concerns included finding sufficient lesson time to deliver the full programme and a lack of time for co-planning and assessment activities.

Despite this, teacher survey and interview data suggest that English Mastery successfully delivered a number of the teacher outcomes from the logic model, including reduction in teacher workload. Although not evidenced by the primary outcome measure, many teachers reported gains in pupil technical accuracy and writing, particularly for those with lower prior attainment or Special Educational Needs and Disabilities (SEND). The embedding of oracy across the programme was seen as beneficial to these pupils. Survey and interview data indicated that most teachers and Mastery Leads enjoyed the training. Programme materials were viewed favourably by most teachers surveyed, particularly for teaching discrete grammar and Writing Mastery. Several teachers interviewed said that Writing Mastery units, combined with the ongoing assessment and tasks during Mastery lessons, helped to scaffold learning for pupils with lower confidence. Improving engagement and motivation is crucial when supporting socioeconomically disadvantaged pupils or those who are behind in English (Ng and Graham, 2018). These findings therefore provide some evidence of a cycle of improvement, with increased confidence through self-efficacy improving pupil motivation (Hempel-Jorgenson et al., 2018), facilitated by the structure of Mastery lessons. Pupils reported increased self-efficacy and motivation in English lessons, particularly Year 8 pupils in schools where the programme was embedded, while teachers reported that pupils benefited from the challenge of the Literary Heritage and Writing Mastery strands. Most teachers perceived improved pupil confidence, corroborating evidence that explicit teaching of vocabulary and proficiency of grammar can support reading comprehension and attainment in writing, key features of effective practice identified in secondary schools by the EEF (EEF, 2021b).

Cost

The average cost per school was around £28,280 over three years, or £78.56 per pupil per year averaged over three years.

Impact

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

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	P Value	EEF cost rating
English attainment	-0.04 (-0.17; 0.08)	0		3780	0.502	£££££
English attainment (FSM pupils)	-0.11 (-0.26; 0.04)	-2	n/a	1112	0.138	£££££

Introduction

Background

English Mastery is designed to improve attainment in English by providing a substantial and ongoing programme of curriculum-integrated continuing professional development (CPD) for teachers. This has the potential to develop knowledge and expertise in English teaching, enabling teachers to deliver a coherent and cumulative Key Stage 3 English experience for pupils where progression is built into the schemes of work and assessment materials of the programme.

Mastery approaches to learning focus on developing and securing pupils' knowledge of key concepts of a subject to support progression and success. The English Mastery approach from Ark Curriculum Plus is described as distinct from other mastery approaches in that it is underpinned by pedagogical pillars: accumulation of knowledge, discrete grammar teaching, Tier 2 vocabulary (words appearing with high frequency in written language but less so in oral language), and use of standardised, norm-referenced assessments. The principles of mastery approaches are delivered through an English curriculum with the following strands.

- *Literary Heritage*: introduces pupils to canonical texts and modern classics. The lessons follow a knowledge-rich approach to teaching English. English literature, the world, history, culture, and heritage are taught, as are connections between these subjects.
- *Writing Mastery*: introduces pupils to grammatical concepts and rules in a logical and cumulative sequence. The rules are taught discretely from reading. The lessons follow an interleaved structure and pupils deliberately practise concepts to the point of mastery.
- *Reading for Pleasure*: introduces pupils to shared reading of contemporary texts using a structured approach. This strand is used to teach vocabulary explicitly.
- Termly pupil assessment.

A comprehensive CPD package is designed to impact on teacher subject knowledge. This includes improving the teaching of canonical texts, grammatical concepts, and vocabulary and understanding of how to use assessment for planning, delivery, and effective feedback. This, in turn, has the potential to impact on pupil vocabulary and knowledge of grammatical concepts and canonical texts as well as on pupil self-efficacy, motivation, and engagement in English lessons (see Figure 1: English Mastery logic model). Research evidence suggests that these pedagogical pillars have the potential to impact on pupil attainment in English. The use of whole texts in English teaching is essential for engagement with texts and reading for pleasure (Cremin, 2014) and the Ark strategy ensures that pupils with weaker reading skills can access abridged texts and study at a 'foundation' level, progressing where appropriate to a 'traditional' level. In addition, the English Mastery approach is underpinned by the principles and practices of metacognition (EEF, 2021a). The following features of effective practice have been identified in secondary schools: the inclusion of targeted vocabulary, reading of complex texts through active engagement and application of existing knowledge, and the breaking down of reading of complex texts and of complex writing tasks (EEF, 2021b).

At the time of writing, English Mastery has been delivered in 328 schools since it was founded in 2014. The EEF previously funded an efficacy trial (Davies, Read and Guscott, 2022), yet this was severely affected by the Covid-19 pandemic and the closure of schools to most pupils from March to September 2020 and again from January to March 2021. Delivery was disrupted as schools were unable to teach in person. As it was not possible to collect the outcome assessment data required for the impact evaluation, this part of the original trial was not completed.

The implementation and process evaluation of the previous English Mastery trial was able to report findings despite these difficulties. It showed that 80% of teachers (n =165) in intervention schools responding to an end of intervention survey believed that English Mastery had reduced their planning and marking workload, 70% perceived that it had improved overall progress in English, and 61% reported that their teaching of classic literature had improved. Most teacher survey respondents (75%) also reported being satisfied with the programme overall, and with the training and support in particular

(84%). Two key outcomes for pupils were increased enjoyment of English and a higher level of writing accuracy. However, some experienced teachers reported that the programme was overly prescriptive, limiting creativity in lesson planning. Adherence to co-planning and formative and summative assessments was also found to be mixed.

Specific changes made to the English Mastery programme since the previous trial include:

- further integration of assessment procedures and the use of these to inform planning;
- improving emphasis and clarity on the importance of co-planning and how it contributes effectively to the high-quality delivery of the programme;
- additional support for teachers in adapting planning to meet the needs of SEND, EAL, and FSM pupils, or those at the lower end of the attainment scale; and
- review of texts included in the programme to include contemporary texts.

A significant feature of the programme is an integrated summative and formative assessment model, referred to throughout this report as 'the assessment protocols'. This involves standardised assessments for each Mastery unit including a comprehensive comparative judgement standardisation booklet covering a range of student performance to run alongside assessing for mastery standardisation days where staff are involved in collaborative assessment and planning. Formative assessment involves frequent mastery quizzes carried out fortnightly.

The implementation and process evaluation of the previous efficacy trial of English Mastery suggested an increase in pupil enjoyment of English (Davies, Read and Guscott, 2022). Literacy engagement involves motivation, decision-making, thinking, effort expenditure, emotion, and the commitment of time. The structured approach to teaching texts in English Mastery has the potential to strengthen pupil cognition and progress in reading and writing and to improve pupil self-efficacy and confidence. Hempel-Jorgenson et al. (2018) suggest a virtuous cycle of increased confidence through self-efficacy that increases pupil reading motivation. Motivation is linked to appropriate goals, perceptions of self-efficacy (Guthrie and Wigfield, 2000), and belonging to peer groups and social relations (Wentzel and Wigfield, 1998; see also DfE, 2012). Improving overall engagement in English has been seen as particularly crucial when supporting pupils who have fallen behind or those from disadvantaged groups who are likely to benefit from teaching that improves self-efficacy, motivation, and engagement (Ng and Graham, 2018; Guthrie et al., 2007).

This trial sought to evaluate English Mastery through a randomised controlled trial to assess the impact of the intervention on pupil attainment in English, the subject on which it is focused. A total of 72 schools were recruited to the study, which adopted a two-arm, three-level design with pupils clustered into classes and classes clustered into schools. The intention was to improve understanding of what works in teaching English for KS3, contributing to the evidence base for a Key Stage which no longer includes statutory assessments. Due to recruitment difficulties, the sample was split into two cohorts, recruited in Spring 2022 and 2023 respectively. The smaller sample size led to a less sensitive trial design, discussed in full in the section on evaluation methods.

Although English Mastery involves training and support for all teachers in participating English departments, the presence of setting and streaming practices in secondary schools informed the decision to incorporate a class level into the study design. The IPE was designed to provide a variety of perspectives on the programme, with teachers taking part in online surveys and, in case study schools, in-person interviews. Lesson observations and pupil focus groups also took place. The intention was to contextualise the impact evaluation findings and generate evidence to examine the logic model.

Intervention

Name

English Mastery.

Why (theory/rationale)

The aim of the English Mastery secondary programme is to improve achievement in Key Stage 3 English. This is done by providing teachers with subject-specific training, curriculum materials, and ongoing in-school support and coaching.

Who (recipients)

Key Stage 3 pupils (for this trial, Years 7 and 8, ages 11 to 13).

What (procedures)

English Mastery is delivered through in-class teaching during regular timetabled English lessons as a whole-curriculum programme. It is a comprehensive programme with two pathways: the traditional curriculum for pupils working at the expected level for their age and the foundation curriculum for those below the age-expected level. If pupils attain the age-expected level in two consecutive assessments conducted as part of intervention delivery then it is recommended that they advance from foundation to traditional.

Specifically, the programme entails induction training for KS3 teachers of English, termly training sessions for the member of staff selected as the Mastery Lead responsible for leading the programme in each school, regular co-planning time, and termly standardised pupil assessments. Classroom teaching of Literary Heritage for 100 minutes per week and Mastery Writing for 50 minutes per week is also expected, while the Mastery Lead is supposed to continue teaching at least two English lessons per week and facilitate school visits and development sessions from the delivery team. Each of these activities is treated as a compliance indicator for this trial. Further detail on measuring compliance can be found in the Methods section.

What (materials)

Training and support to develop teacher confidence and expertise to deliver the programme involved a teacher/leader development programme and a curriculum and assessment programme (see Figure 1). These two elements aimed to develop confidence in the links between the planning and assessment cycle (weekly co-planning of lesson delivery and design, units of work, lesson-by-lesson resources, and integrated assessment model) and the teacher development programme (induction training, school support sessions and visits, teacher development videos and workshops, and Mastery Lead training).

Weekly departmental co-planning materials are provided to aid teachers in adapting the English Mastery lessons for their pupils. Teachers also receive training and support through induction, additional training for the KS3 lead, remote development sessions, biannual school coaching visits, 'Assessing for Mastery' standardisation sessions each term, and optional webinar workshops. Pupil assessments, undertaken termly, are central to the intervention. Teachers have the chance to compare assessments with those from other schools. All teachers receive access to the MyMastery platform with the full range of curriculum resources and on-demand training materials, classroom footage, and training events.

All pupils receive the same dosage and study the same topics but the foundation curriculum is designed to be accessible to lower-attaining pupils (such as through use of abridged texts). The foundation curriculum includes the use of abridged texts to provide access for weaker readers. Teachers are encouraged to adapt both the foundation and traditional curriculum teaching materials to meet the needs of individual pupils during lessons. Strategies to support a diverse range of pupils' needs are shared by the developers as part of ongoing training sessions and via the MyMastery platform.

Who (provider)

The intervention is designed by Ark Curriculum Plus, which also runs the training and provides all ongoing support to the teachers delivering English Mastery in schools.

How (format)

Schools recruited to the trial and randomised into the intervention group send all KS3 English teachers for external training provided by the developer. A member of the English department must be nominated as Mastery Lead to lead the implementation of the programme in their school. They lead weekly co-planning sessions to support teachers to plan their implementation of the curriculum together. Teachers then deliver English Mastery through normal timetabled English lessons in school. This intervention therefore targets the entire pupil cohort (for this trial, Year 7 during the first year of delivery and Year 8 during the second) in participating schools.

Where (location)

The intervention is delivered in classrooms. Schools in England with KS3 pupils studying English were eligible for the trial. This project was part of the DfE's Accelerator Fund and there was a requirement for a minimum of 50% to be recruited from three Regional School Commissioner (RSC) regions of England: West Midlands, East Midlands and the Humber, and the North.¹ Teacher induction training is provided offsite. Ongoing support is provided online or in school.

When and how much (dosage)

The sample was split into two cohorts, recruited in Spring 2022 and 2023 respectively. This was due to difficulties in recruiting the intended number of schools, as discussed below. For the first cohort, KS3 English teachers at schools randomly allocated to the intervention group attended English Mastery training in September 2022. English Mastery was then delivered to pupils through their usual timetabled English lessons during the 2022/2023 and 2023/2024 school years. The 2022/2023 Year 7 cohort continued with the programme into Year 8. The second cohort followed the same timetable starting one year later, with two years of delivery beginning in Autumn 2023. The programme covers all of Key Stage 3, from Year 7 to Year 9, but for this trial the evaluation sample of pupils was taught in Year 7 and Year 8. The curriculum design is cumulative, with content from previous terms and year groups being revisited and expanded in future years. Similarly, the teacher development programme is also cumulative, developing pedagogical, subject, and assessment knowledge across multiple years.

Tailoring (adaptation)

The intervention is designed for all KS3 (Year 7 to Year 9), however, due to funding constraints the current trial only included Years 7 and 8—this applies to both recruitment cohorts. While Ark provides comprehensive training and resources to deliver English Mastery, teachers are given 'creative freedom' when planning lessons and teachers are invited to customise the materials to suit each class to address the 'full spectrum' of needs. How teachers adapted the programme and resources is discussed in further detail in the IPE section.

Control condition

No specific data was gathered to identify business as usual in intervention or control schools. However, school motivations for joining the programme can provide insights on this. Schools said they had joined the programme as they wished to be more confident that their English curriculum had the required consistency, progression, and coverage and provided strong preparation for GCSE English. Some leaders wanted to take pressure off staff in planning and assessing in English and teacher subject knowledge, particularly for non-specialist English teachers, was a concern. These comments expressed during interviews suggest that school leaders perceived that business as usual was not meeting some of these criteria.

Control schools taught KS3 English under business-as-usual conditions during the study period. Incentive payments totalling £1,250 were paid to control schools in three instalments: £500 for submitting pupil and class data (Autumn 2022

¹ As defined by Regional Schools Commissioner boundaries (see [here](#) for local authorities included).

for Cohort 1; Autumn 2023 for Cohort 2), £250 for fulfilling all evaluation requirements by the end of the first year, and £500 for completing the pupil outcome assessment (Summer 2024 for Cohort 1; Summer 2025 for Cohort 2). Control schools were not able to purchase English Mastery on a commercial basis during the trial.

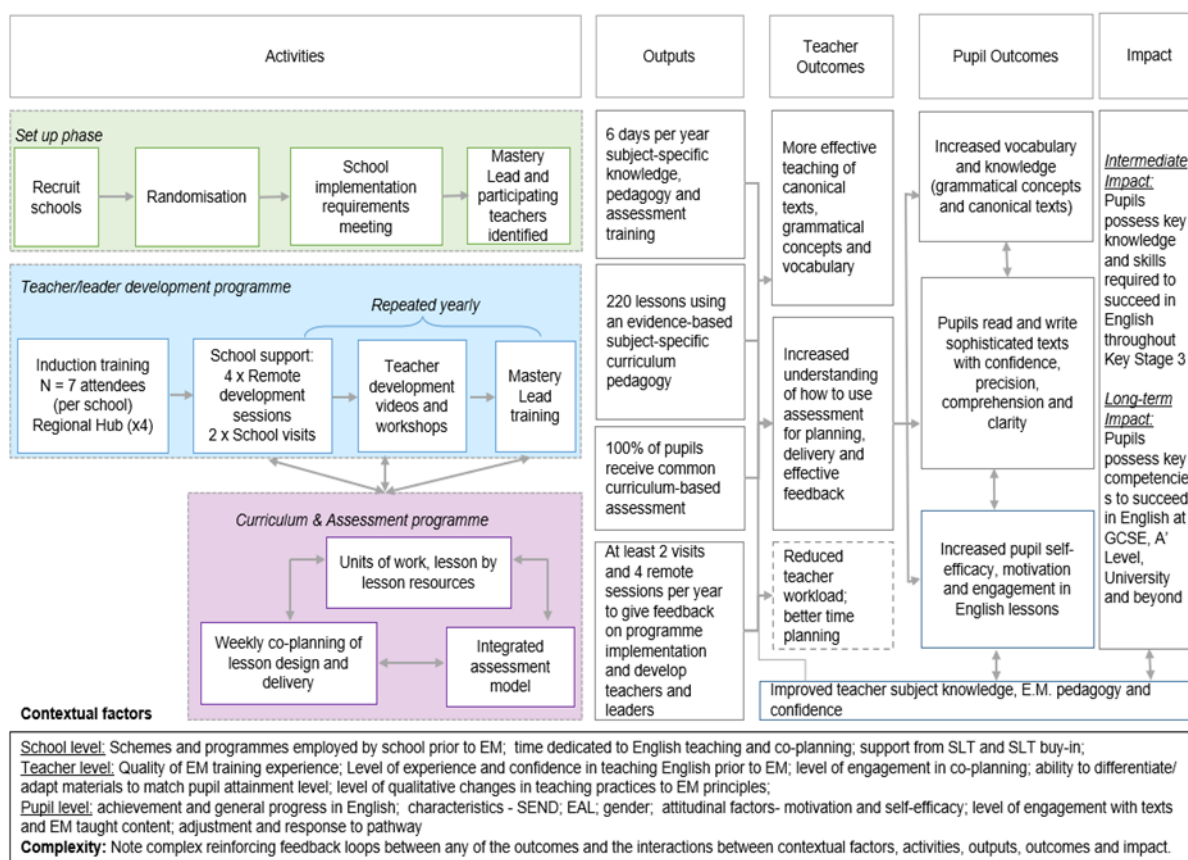
Theory of change and logic model

The Ark Curriculum Plus English Mastery approach is designed to improve attainment in English by providing a substantial and ongoing programme of support and curriculum-integrated professional development for teachers. This has the potential to strengthen knowledge and expertise, enabling teachers to deliver a coherent and cumulative Key Stage 3 English experience for pupils.

In defining a theory of change, we draw on the work of Maxwell and Coldwell (2018, p. 269) who follow Merton in conceptualising theories as those ‘that lie between the minor but necessary working hypotheses [...] and the all-inclusive systematic efforts to develop a unified theory’ (1968, p. 39). The aim is to understand ‘the mechanisms that mediate between the delivery (and receipt) of the program and the emergence of the outcomes of interest’ (Weiss, 1998, p. 57). These mechanisms are detailed in the logic model.

The logic model includes self-efficacy, motivation, and engagement as pupil outcomes. The relationship between increased self-efficacy and motivation is noted in the literature (Hempel-Jorgenson et al., 2018). However, the previous trial used the term ‘enjoyment’, hence its inclusion in the research questions (see page 11). This also aligns with decision to incorporate pupil perspectives into the research design.

Figure 1: English Mastery logic model



Evaluation objectives and research questions

Impact evaluation research questions

Primary research question

1. What is the impact of English Mastery on the English attainment of Year 8 pupils, as measured by the GL Progress Test in English (PTE)?

Secondary research questions

2. What is the impact of English Mastery on two dimensions of English (Spelling, Grammar and Punctuation and Reading Comprehension) measured by the GL PTE?
3. What is the impact of English Mastery on the English attainment of Year 8 FSM pupils, as measured by the GL PTE?
4. What is the impact of English Mastery on the English attainment of Year 8 pupils with lower prior attainment, EAL, and SEN as measured by the GL PTE?

Implementation and process evaluation research questions

- a) Are any changes in self-reported teacher workload associated with English Mastery?
- b) To what extent does the English Mastery training, its scheduling, and materials equip teachers and leaders to implement a coherent and cumulative English curriculum?
- c) In what ways has English Mastery led to changes in practice in the teaching of English, including for supporting the needs of different pupil groups (such as EAL, SEND, and FSM and those on traditional/foundation pathways)?
- d) How have the English Mastery curriculum materials led to changes in the teaching of English?
- e) Have English Mastery assessment protocols led to changes in teacher use of formative and summative assessment to inform planning and teaching?
- f) What are teacher perceptions of the impact of English Mastery on the attainment and enjoyment of English for different groups of pupils?
- g) What are pupil perceptions of the impact of English Mastery on their self-efficacy, confidence, and enjoyment in English?
- h) What are school perceptions of the programme representing value for money?

The Statistical Analysis Plan (**SAP**) and trial **protocol** are both published on the EEF website.

Ethics and trial registration

The evaluation was approved by the SHU ethics committee on 26 January 2022 (Ethic Review ID: ER40773698).

The trial has been publicly registered on ISRCTN: <https://www.isrctn.com/ISRCTN76965629>

Schools signed a Memorandum of Understanding (MoU) to confirm their place on the trial. The MoU for both delivery cohorts can be found in Appendix D (see Additional Appendices). Schools received a participant information sheet to be sent to the parents/carers of all pupils taking part in the trial before their data was shared with the evaluation team. This was sent to schools once the pupils started Year 7 at participating schools (in September 2022 for Cohort 1, September 2023 for Cohort 2). It set out the details of the data processing and included a slip to be returned to the school for parents/carers who wished to withdraw their child from the evaluation (see Appendix D in Additional Appendices).

Data protection

This EEF trial was part of a wider Department for Education-funded programme called the 'Accelerator Fund'. SHU was the data controller for the English Mastery evaluation and the processing of personal data is defined under GDPR as a specific

task in the public interest. As data was processed for the purpose of academic research, the main aim of which is to improve reading ability among school pupils, the legal basis for processing is as a 'Public Task' (Article 6 (1) (e)): see <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/public-task/>

Special category data, specifically English as an Additional Language (EAL), Special Educational Needs and Disabilities (SEND), and Free School Meals (FSM) status, was accessed from the National Pupil Database and processed for the purpose of scientific research as permitted under GDPR Article 9 (j). Pupil names, dates of birth and Unique Pupil Numbers (UPNs) obtained from schools were used to access this information: see <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category-data/>

Specifically, this data was processed to determine whether the English Mastery programme has different effects on different subgroups of pupils. The EEF was established with a remit to break the link between family background and educational attainment and all EEF projects conduct subgroup analyses on FSM pupils. The effectiveness of the English Mastery programme for pupils who have EAL and SEND were also of interest. Further details can be found in the privacy notice (also included in Appendix D): <https://www.shu.ac.uk/-/media/home/research/sioe-rke/privacy-notices/eef-english-mastery-privacy-notice.docx>

Project teams

Evaluation team—Sheffield Hallam University

Dr Martin Culliney: principal investigator and impact evaluation lead.

Dr Karen Daniels: co-investigator and IPE lead.

Sean Demack: impact evaluation.

Lewis Clark : impact evaluation and IPE.

Dr Marie Helks: IPE.

Martin Illingworth (formerly Sheffield Hallam University): IPE.

Delivery team—Ark Curriculum Plus

Nick Wallace: project lead.

Michlyn Caffrey: head of English Mastery Secondary.

Kirsty Kelsey: programme design and delivery.

Formerly Ark Curriculum Plus:

Dr Sarah Dowey: programme delivery.

Caroline Redgate: programme delivery.

Kathy Watson: programme delivery.

Stephanie Keenan: programme design and delivery.

Hope Vardon-Prince: programme delivery.

Lyndsay Bawden: programme delivery.

Sarah Haigh: programme delivery.

Ginny Kearns: project management.

Megan Latchford: project management.

Methods

Trial design

Table 3: Trial design

Trial design, including number of arms		Two-arm, three-level cluster randomised.
Unit of randomisation		School.
Stratification variable(s) (if applicable)		Geographical area (Regional School Commissioner regions); setting/streaming practice; OFSTED rating.
Primary outcome	Variable	English attainment.
	Measure (instrument, scale, source)	GL Progress Test in English total raw score (0–65 scale).
Secondary outcome(s)	Variable(s)	Reading; spelling, punctuation and grammar.
	Measure(s) (instrument, scale, source)	GL Progress Test in English (Reading: 0–32 scale; Spelling, Punctuation and Grammar: 0–33 scale).
Baseline for primary outcome	Variable	English attainment.
	Measure (instrument, scale, source)	KS2 English from NPD (KS2 Reading: 0–50 scale; KS2 GPS: 0–70 scale, combined).
Baseline for secondary outcome(s)	Variable	English attainment.
	Measure (instrument, scale, source)	KS2 Reading and GPS from NPD (as for primary outcome but separate scales).

This two-arm, three-level clustered efficacy trial recruited and randomised schools in two cohorts (Year 7 in 2022/2023 and 2023/2024). Pupils were clustered into classes and classes clustered into schools. The primary outcome was English attainment as measured by the total raw score on the GL Progress Test in English 13, and the secondary outcomes were subscales from this test (Reading; Spelling, Punctuation, and Grammar). Outcome measures are discussed in full below.

The original plan was to have only one cohort, but recruitment difficulties in Spring 2022 meant that more schools were needed to reach the target sample size. Delivery for each cohort lasted for two school years (September 2022 to July 2024 for Cohort 1, September 2023 to July 2025 for Cohort 2), from when pupils started Year 7 until they finished Year 8. Control schools continued with a business-as-usual approach during the trial. There was no waitlist or alternative treatment. As mentioned above, control schools completing all evaluation activities received a total of £1,250 in payments for their participation.

Participant selection

The target was to recruit 100 schools with pupils in Year 7 and Year 8 during the evaluation period, with at least 50% based in either the West Midlands, the North of England, or the East Midlands and the Humber regions. In the end, 72 schools signed up, although 45 (62.5%) were not from the target regions due to difficulties recruiting in those areas.

Recruitment was led by the developer. Activities to promote the trial included speaking at national events, marketing campaigns over email and social media, and webinar events. Information was also shared through the developer's existing contacts in the education sector, EEF distribution lists, and the Research School network.

It was agreed that schools partaking in the previous trial as members of the control group would be eligible for this trial if they had no exposure to English Mastery and were willing to accept the outcome of randomisation, although ultimately no such schools were recruited. For further details about participating schools, see Tables 11 and 12 (below).

Outcome measures

Baseline measures

Key Stage 2 English attainment was used as the baseline covariate for analyses of the primary outcome. The KS2 Reading (KS2_READMRK) and Grammar, Punctuation and Spelling (KS2_GPSMRK) raw scores were standardised (with the sample mean subtracted from each value, which was then divided by the standard deviation) and summed to create a single combined baseline measure for use as a covariate in analysis of the primary outcome. This was to ensure that both test scores contributed equally to the overall measure of prior attainment in English. For the secondary outcomes, the separate KS2 Reading and GPS scores were used as baseline measures. Following disruption to statutory assessments in 2020 and 2021 due to the Covid-19 pandemic, testing resumed in 2022 and it was possible to use KS2 test results from NPD as intended. This was preferred to the alternative of using a commercial test such as GL PTE at baseline as administrative data ordinarily provides higher response rates and is more cost-effective while reducing burden on schools. Correlations between the baseline and endline outcome measures used are presented below (Table 9).

Primary outcome

The primary outcome is the total raw score on the GL PTE132, taken by Year 8 pupils in summer 2024 (Cohort 1) or summer 2025 (Cohort 2). This assessment was selected as the content was deemed appropriate for the pupil outcomes targeted by the intervention, having been chosen as the primary outcome in the previous efficacy trial of English Mastery. The paper version was used for ease of administration, with schools told to allow one hour for completion. Schools received detailed instructions from the evaluation team advising that the paper must be undertaken under test conditions, either in class or with all pupils sat together in an exam hall. Completed papers were sent back to the evaluation team, who posted them to GL Assessment for marking.

Secondary outcomes

Subscales from the GL PTE13 test were used as secondary outcomes, as in the previous trial of English Mastery. Specifically, four subscales are provided in GL PTE13, and three are relevant here. First, Spelling, Punctuation, and Grammar, measured on a 0–33 scale, was analysed as a secondary outcome with KS2 Grammar, Punctuation, and Spelling used as the baseline measure. The two reading scales comprising GL PTE13—Reading Comprehension Narrative and Reading Comprehension Non-Narrative—were summed to form an aggregated outcome measure of reading (scale 0–32) for analysis alongside the baseline measure KS2 Reading.

Sample size

The recruitment target of 100 schools was not met by the original deadline in summer 2022, with only 39 schools recruited. It was agreed that these schools should start the trial as intended, with intervention schools taking part in English Mastery

² <https://www.gl-assessment.co.uk/products/progress-test-in-english/>

during the 2022/2023 and 2023/2024 school years. A second cohort was then recruited in 2022/2023 with a further 33 schools agreeing to take part in the study during 2023/2024 and 2024/2025. In total, 72 schools were recruited across the two study cohorts. The power calculations for the randomisation stage presented in Table 9 are based on this number. It was decided to select a sample of pupils from within each class to reduce testing costs as this strengthens the power calculations by increasing the number of level two clusters without increasing the number of pupils. Furthermore, as some schools use setting or streaming, sampling certain classes would heighten the risk of biasing the selection toward the upper or lower ends of the attainment distribution.

Statistical sensitivity is commonly illustrated using a Minimum Detectable Effect Size (MDES), the smallest effect detectable by a proposed RCT design as statistically significant (often set as $p < 0.05$) with a statistical power of 80% or higher. Equation 1, drawn from Kelcey et al. (2017), was used to estimate the MDES for this trial design and the figures were checked with PowerUp! (Dong et al., 2015, sheet BCRA4_3f). Estimates are based on previous research into class-level clustering (Demack, 2019) and evaluations of KS3 maths (Boylan et al., 2015; Demack et al., 2022).

Equation 1: MDES calculations

$$MDES_{3LCRT} \sim M_{K-L-2} \sqrt{\frac{1}{P(1-P)}} \sqrt{\frac{ICC_{sch}(1-R_{sch}^2)}{K} + \frac{ICC_{class}(1-R_{class}^2)}{JK} + \frac{(1-ICC_{sch}-ICC_{class})(1-R_{pup}^2)}{nJK}}$$

For this three-level clustered design, the MDES is influenced by:

- n = number of pupils per class, allowed to vary between three (estimated number of pupils eligible for FSM) and 12 (whole sample); J = number of classes per school (fixed at six); K = number of schools (72);
- P = proportion of schools allocated to intervention group (= 0.5 at protocol; 0.45 at randomisation);
- L = number of (level three) covariates used (six: group membership, school OFSTED rating, setting/streaming and three geographical region dummies);
- M_{K-L-2} is the group effect multiplier value of the t-distribution for a two-tailed test with $\alpha = 0.05$ and $\beta = 0.80$, equal variances assumed with $(K-L-2) \sim (72-6-2 = 64)$ degrees of freedom M_{64} for 72 schools (or M_{57} for 65 schools allowing 10% attrition);
- ICC_3 is the school-level Intraclass Correlation Coefficient (proportion of variance at level three) ~ unknown but estimated at 0.10 (from first efficacy trial);
- ICC_2 is the class-level ICC (proportion of variance at level two) ~ unknown but estimated conservatively as 0.40 with the assumption that setting/streaming in KS3 English will be common across recruited schools;
- R_{pup}^2 = proportion of within-class pupil-level variance that is reduced by covariate(s) - pupil level explanatory power ($R_1^2 = 0.49$ based on a correlation of 0.70);
- R_{class}^2 = proportion of between classes within-schools variance that is reduced by covariate (prior English attainment) - class level explanatory power ($R_2^2 = 0.81$ based on a correlation of 0.90), slightly smaller than observed in Demack et al. (2022) in a study of secondary maths as it was assumed that setting/streaming is slightly less common in KS3 English than in KS3 maths; and
- R_{sch}^2 = proportion of between-schools variance that is reduced by covariate(s) – school-level explanatory power – ($R_3^2 = 0.16$ based on a correlation of 0.40), estimated more conservatively than the previous English Mastery trial.

With a total of 72 schools (as randomised), the outcome testing would have included 5,184 pupils (1,296 FSM pupils) in 432 classes and the resulting MDES estimates are 0.21 (overall) and 0.22 (FSM). These figures are reported below in Table 9. The MDES does not improve with more than 12 pupils per class, which justifies the sampling approach as discussed below.

Randomisation

Researchers in the evaluation team allocated schools to the intervention or control group through minimisation using the MinimPy software. This aimed to balance the two treatment groups on the three key variables: geographical area (West Midlands, North, East Midlands and Humber, other non-Accelerator Fund regions), setting/streaming practices ('yes' or 'no'), and OFSTED rating (outstanding/good or requires improvement/inadequate). It also allowed additional schools to be allocated in the second cohort while accounting for the characteristics of the first cohort.

Geographical region was included to produce a more even spread of schools across the participating areas. Setting and streaming practices are important due to the expected effect on class-level clustering in the outcome data. OFSTED rating has been used as an indicator of school quality as attainment data was not available for the two years prior to the trial.

Schools were assigned a unique random number prior to allocation. For Cohort 1, the 39 participating schools were each allocated a random number between 1 and 39; for Cohort 2, the 33 schools were each allocated a random number between 1 and 33. These numbers determined the sequence in which schools were entered into MinimPy for allocation.

Within the first cohort, schools that had signed the MoU before 22 June 2022 were allocated to either the intervention or control group in two batches on that date; a second batch was allocated on 13 July 2022. The second cohort was recruited during the Spring term 2023 and allocated in two batches: in May 2023 and July 2023. Once all schools within a cohort had been allocated to either the intervention or control group, schools were notified of their allocation by the delivery team.

Analysis was not blinded to randomisation. The evaluation team was aware of group allocation as this was necessary for the administration of surveys and the coordination of subsequent fieldwork with participating schools. To mitigate potential bias associated with unblinded analysis, the evaluation team had no contact with recruited schools prior to allocation. In addition, all primary analyses were pre-specified in the trial protocol and SAP prior to randomisation, and both baseline and outcome assessments were externally marked.

The following data was collected from schools before randomisation: contact details for school contacts and English teachers, whether the school uses setting/streaming in KS3 English, the number of timetabled hours for English each week, and information about prior use of approaches similar to English Mastery. After meeting these requirements, schools were included in the allocation process.

Schools undertook to provide the full name, date of birth, and unique pupil number for each pupil in the 2022/2023 Year 7 cohort at the start of the Autumn term so that KS2 test results and pupil characteristics could be obtained from the NPD. Schools were required to specify the class and teacher for each pupil and provide an update on this in 2023/2024. For the second cohort, Year 7 data was submitted in Autumn 2023/2024 with the Year 8 update in 2024/2025.

Statistical analysis

Primary analysis

Given the nested structure of the data, a multilevel approach was adopted as per EEF statistical analysis guidance, with pupils clustered into classes and classes clustered into schools. Multilevel linear regression models were constructed for the GL PTE13 primary outcome. KS2 English attainment (Reading and GPS raw scores standardised and combined) was used as the baseline covariate for analyses of the primary outcome. The first model included only the school-level group identifier. This model produced the unconditional variance used as the denominator for the effect size calculations.

The second model added KS2 English as a covariate at the pupil, class, and school levels and was only used to examine variance decomposition as reported below in Table 18. Prior attainment covariates were centred so that the school level was centred on the mean for each school (the grand mean was subtracted from the school mean); the class level was centred around the school mean; the pupil level was centred around the class mean. These prior attainment covariates were included in the analysis to improve estimate precision by accounting for compositional effects.

The final model added variables used in the minimisation (geographic area, setting/streaming indicator, and OFSTED rating) and formed the headline intention-to-treat (ITT) model for the primary outcome. It is represented by the following equation:

Equation 2: Headline analysis model

$$Y_{ijk} = b_0 + b_1 \text{Group}_k + b_2 \text{Baseline}_{ijk} + b_3 \text{Baseline_class}_{jk} + b_4 \text{Baseline_school}_k + b_5 \text{Sets_streams}_k + b_6 \text{OFSTED}_k + b_7 \text{EastMidsHumber}_k + b_8 \text{North}_k + b_9 \text{WestMids}_k + u_k + r_{jk} + e_{ijk}$$

—where Y_{ijk} is the outcome for pupil i in class j in school k , b_0 is the constant, and Group is a binary indicator of school treatment allocation. Pupil-, class-, and school-level prior attainment covariates are represented by Baseline , Baseline_class , and Baseline_school . The stratifiers used in the randomisation are denoted as (a) Sets_streams , a binary indicator of whether the school uses setting/streaming or teaches mixed ability groups, (b) OFSTED , a binary indicator of whether the school is rated as good or outstanding by OFSTED as opposed to other ratings, and (c) three region dummies with all other schools as the reference category. The random errors are represented by u_k (school level), r_{jk} (class level), and e_{ijk} (pupil level). For each outcome measure, the coefficient of the school-level group variable ($b_1 \text{Group}$) was converted into Hedges' g effect size statistics with 95% confidence intervals using the unconditional variance from the outcome only model described above.

Secondary analysis

For each secondary outcome measure, the analysis followed the same procedure as for the primary outcome. The baseline measures used in these models were detailed above.

Analysis in the presence of non-compliance

Compliance was assessed at the school level. The compliance measures specified in Table 4 were adapted from those used in the previous English Mastery trial. During the course of this trial, minor changes have been made to item two, and item seven has been added to reflect developments in the expectations on participating teachers. This was at the developer's request, intended to reflect changes to the delivery of the programme over the evaluation period. These changes were specified in the statistical analysis plan and the most recent version of the trial protocol. Information on these compliance indicators was collected by the developer in the course of its work with intervention schools and shared with the evaluators at the end of the trial. Compliance was defined as a binary variable and to be classed as compliant, schools had to meet the threshold for all indicators.

Table 4: Compliance indicators and thresholds

Item	Compliance indicator	Overall threshold
1. Induction training attendance	KS3 teachers attend EM induction training.	80% of teachers delivering EM attend induction training.
2. Mastery Lead training attendance	EM Lead attends a termly training session (Mastery Lead workshop or Assessing for Mastery).	Attendance in at least 2/3 terms in Year 1 and in Year 2.
3. Co-planning time allocation	Co-planning allocated fortnightly at least ten weeks per term.	At least 2/3 terms in Year 1 and in Year 2.
4. Standardised conditions for assessment	Pupils sit the termly EM assessments in standardised conditions.	At least 2/3 terms in Year 1 and in Year 2.
5. English Mastery curriculum delivery	Schools teach Literary Heritage for 100 mins per week and Mastery Writing for 50 mins per week for at least ten weeks per term.	At least 2/3 terms in Year 1 and in Year 2.
6. English Mastery Lead involvement in teaching	EM Lead teaches at least two lessons per week for a minimum of ten weeks per term.	The English Mastery Lead teaches two lessons per week for 27 weeks of the year (70%).
7. Support uptake	EM Lead facilitates touchpoints with their school.	Four touchpoints each year: two school visits and two development sessions.

These variables were used to estimate the Complier Average Causal Effect (CACE). The purpose of the CACE analysis is to estimate the impact of English Mastery for pupils deemed to have ‘complied’ with the intervention. In this case, this refers to all the pupils attending schools that were classified as compliers. CACE is estimated using two-stage least squares (2SLS) regression. The evaluation SAP specified that predicted compliance would be estimated using multilevel logistic regression, yet the decision was taken to use the `ivregress` command in Stata 17 to explore whether compliance was related to outcome scores, and whether compliance was endogenous, following the approach used in Anders et al. (2021). The two-stage equation first models compliance against the other covariates included in the headline model:

Equation 3: Two-stage least squares model—first stage

$$\text{Compliance}_{ijk} = b_0 + b_1\text{Group}_k + b_2\text{Baseline}_{ijk} + b_3\text{Baseline_class}_{jk} + b_4\text{Baseline_school}_k + b_5\text{Sets_streams}_k \\ + b_6\text{OFSTED}_k + b_7\text{EastMidsHumber}_k + b_8\text{North}_k + b_9\text{WestMids}_k + e_{ijk}$$

The second stage then adds the compliance variable to the headline model of the primary outcome in place of the treatment allocation indicator:

Equation 4: Two-stage least squares model—second stage

$$Y_{ijk} = b_0 + b_1\text{Compliance}_k + b_2\text{Baseline}_{ijk} + b_3\text{Baseline_class}_{jk} + b_4\text{Baseline_school}_k + b_5\text{Sets_streams}_k \\ + b_6\text{OFSTED}_k + b_7\text{EastMidsHumber}_k + b_8\text{North}_k + b_9\text{WestMids}_k + e_{ijk}$$

Missing data analysis

The impact analyses examined missing data in the outcome variables and considered whether it was reasonable to assume that the missing data was random. A multilevel logistic regression model with a binary outcome identifying when outcome data is missing (=1) or not (=0) and the same covariates as the headline ITT model was estimated to examine any patterns. The SAP stated that if over 5% of pupil outcome data is missing, a multilevel logistic regression model with a binary outcome identifying when outcome data is missing would be constructed and that the ITT variables and additional school-level variables would be used to identify whether the missing outcome data can be assumed to be missing completely at random. If none of the explanatory variables were found to account for a statistically significant amount of variation in the missing outcome data, we would cautiously assume that the data is missing completely at random (MCAR), otherwise multiple imputation will be used and the results compared with the headline ITT analysis for the primary outcome.

Subgroup analyses

Follow-on ITT analyses focused on the impact of the English Mastery programme on English attainment for pupils eligible for FSM, identified by the FSM_6_P indicator from the NPD. The same approach used for the headline ITT analyses was repeated on the FSM subgroup and on the whole ITT sample with an interaction effect for treatment allocation and FSM status as per EEF analysis guidance. The effect sizes for both models are presented in the Results section of this report.

Further exploratory analysis was undertaken for the following subsamples: the two delivery cohorts, pupils with low prior attainment in KS2 English, pupils classed as EAL in NPD data, and pupils classed as SEN in NPD data. The trial is not powered for any of these analyses yet the effect sizes are presented in case the results differ noticeably from those estimated for the ITT sample.

Additional analyses and robustness checks

The intervention was delivered to two separate recruitment cohorts as described above. It was decided that replicating the headline ITT model on each cohort separately would be a worthwhile sensitivity analysis. Results are reported in the Impact Evaluation section.

Estimation of effect sizes

The effect size measure used is Hedges' g . This was calculated using the following equation.

Equation 5: Effect size calculation

$$ES = \frac{(T - C)_{adjusted}}{\sqrt{\delta_{sch}^2 + \delta_{class}^2 + \delta_{pup}^2}}$$

where:

- δ_{sch}^2 is the school-level variance, δ_{class}^2 is the class-level variance, and δ_{pup}^2 is the pupil-level variance for the PTE13 outcome from the outcome only multilevel model; and
- $(T - C)_{adjusted}$ is the mean difference between the attainment of pupils in treatment schools and pupils in control schools; this is obtained from the coefficient for the school-level 'group' variable from the headline ITT analyses (b_{Group} in Equation 2 above); the coefficient standard error and the upper/lower 95% confidence intervals were also converted into units of standard deviations using the above formula.

Estimation of ICC

For both pre and post-test, ICCs at the school level were estimated using an outcome only 3-level multilevel variance components model. In the results section, the variance decomposition for the three levels (school, class, and pupil) along with the ICC estimates at school and class level are presented.

Equation 6: ICC calculation, school level

$$ICC = \frac{\text{Variance}_{school}}{\text{Variance}_{school} + \text{Variance}_{class} + \text{Variance}_{pupil}}$$

Equation 7: ICC calculation, class level

$$ICC = \frac{\text{Variance}_{school} + \text{Variance}_{class}}{\text{Variance}_{school} + \text{Variance}_{class} + \text{Variance}_{pupil}}$$

Implementation and process evaluation

The implementation and process evaluation set out to answer the following research questions.

- Are any changes in self-reported teacher workload associated with English Mastery?
- To what extent does the English Mastery training, its scheduling, and materials equip teachers and leaders to implement a coherent and cumulative English curriculum?
- In what ways has English Mastery led to changes in practice in the teaching of English, including for supporting the needs of different pupil groups (such as EAL, SEND, and FSM and those on traditional/foundation pathways)?
- How have the English Mastery curriculum materials led to changes in the teaching of English?
- Have English Mastery assessment protocols led to changes in teacher use of formative and summative assessment to inform planning and teaching?
- What are teacher perceptions of the impact of English Mastery on the attainment and enjoyment of English for different groups of pupils?

- g) What are pupil perceptions of the impact of English Mastery on their self-efficacy, confidence, and enjoyment in English?
- h) What are school perceptions of the programme representing value for money?

Research methods

1. Evidence review to identify relevant background literature and discussion with the developers to agree an evidence-informed logic model and data collection methods, taking place at the IDEA workshop where the logic model and its contextual factors were identified.
2. Observations of training events, conducted by experienced literacy specialists, focused on engagement, alignment with expected content, and process. Attendance at training events and other engagement with English Mastery support were examined to gauge fidelity.
3. School visits to understand fidelity, influences on implementation, and the extent to which practice aligns with expectations. Ten schools were each visited once during Year 7 and nine (due to one school withdrawing from the intervention) were visited again in Year 8 to examine perceived impacts of the intervention across the study period. These visits were divided equally between the first and second cohorts of schools and took place in the spring and summer terms to allow enough time for the intervention to become embedded. Schools were selected purposively with a view to visiting schools of varying size and characteristics in terms of variables used in the treatment allocation process (setting and streaming practices, OFSTED rating, geographical region). During each school visit, at least one English Mastery session was observed. This was followed by an interview with the teacher(s) leading the lessons and a focus group of pupils selected by the school, whose lesson was also observed. Observations were guided by a schedule designed to reflect the principles of English Mastery. The Mastery Lead in school was also interviewed. Interview schedules were guided by the logic model.
4. Pre- and post-intervention teacher survey examining understanding and practices around teaching English. The post-intervention survey addressed contextual issues raised in the school visits and from the delivery team. The survey also examined changes in classroom practice associated with English Mastery, with a focus on the 'pedagogical pillars' of the programme, teachers' perceptions of their confidence in a mastery approach, and impact on workload.

The research instruments can be found in Appendix H (see Additional Appendices).

Analysis

Pre- and post-intervention surveys

Survey links were sent to the English Mastery Lead at each intervention school who was then responsible for forwarding the survey to the relevant teachers. In total, this resulted in 763 responses across the trial, however, response rates varied considerably across questions and timepoints. For example, for question 11 in the end of Year 7 teacher survey, the number of responses ranged from 61 to 80 across the eight constituent items.

To mitigate the risk that schools with higher response rates would disproportionately influence the findings, teacher responses were aggregated at the school level (see Table 5). This approach was considered appropriate given that many survey items had a clear school-level element. For example, responses to questions relating to teacher workload are likely to be shaped by school policies and practices, such as timetabling arrangements, which influence individual teachers' experiences and perceptions. While this inevitably concealed some within-school variation in schools where multiple teachers answered survey questions, it was deemed preferable to allowing schools with more responses to distort the sample. Aggregation of the data resulted in 22 out of 35 intervention schools and 26 out of 37 control schools being represented at baseline, 21 intervention schools and 29 control schools at the end of Year 7, and 20 intervention schools and 27 control schools at the end of Year 8.

Table 5: Teacher survey response rate aggregated by school

	Intervention N	Control N	Total N
Baseline	22/35	26/37	48/72
End of Year 7	21/35	29/37	50/72
End of Year 8	20/35	27/37	47/72

As IPE research questions A and C explore change over time across both intervention and control schools, the analysis for these questions was restricted to schools responding to all three surveys (baseline, end of Year 7, and end of Year 8). This resulted in a sample of 13 intervention schools and 24 control schools (see Table 6). In total, 61 teachers across 31 schools responded to all three surveys. In 17 of those schools more than one teacher responded to all three surveys. Only teachers that responded to at least two of the three surveys were analysed. In certain cases, aggregated school responses came from different teachers (or combinations of teachers) at different time points. It was decided that limiting the analysis to teachers that responded to all three surveys would have been too restrictive given that responses were aggregated to school level and some questions were not featured in all three surveys.

Table 6: Teacher survey response rate aggregated by schools represented at all three timepoints

	Intervention N	Control N	Total N
Baseline	13/35	24/37	37/72
End of Year 7	13/35	24/37	37/72
End of Year 8	13/35	24/37	37/72

Descriptive statistics were used to summarise teacher experiences of the programme and other issues relevant to the research questions such as teacher workload. Analysis included school-level pre/post-intervention comparisons. Comparisons between the intervention and control groups were also conducted where possible.

School visits

Five schools' Cohort 1 were visited in Year 7. Four of these were revisited in Year 8. One school withdrew from the programme prior to the follow up visit. Monitoring of Cohort 2 involved visits to five schools in Year 7 and again in Year 8. Across all schools ten leads and ten teachers were interviewed. Each visit involved observing one English Mastery lesson to ascertain how far observations and materials reflected the features of effective pedagogy as described by the developer and the principles underpinning the English Mastery Programme. Focus groups were held during each visit with pupils who had just taken part in an English Mastery lesson. Group size varied from four to seven. All pupil focus groups and interviews were transcribed. Inductive coding was conducted using NVivo software to enable a cross-case analysis to highlight patterns and recurring themes (Miles et al., 2019). As the fieldwork took place over a two-year period for each school—involving Year 7 and Year 8 pupils, teachers, and leaders over the first and second year of implementation—we utilised a variable-orientated approach to case analysis. This enabled us to consider which year group and which year of implementation the data referred to. This enabled us to consider participant perceptions of possible causal relationships and mechanisms of change associated with the logic model and take account of variables that arose during interviews over the two year period. As this analysis was conducted during the first and second year of delivery for both cohorts, cross-case themes were compared to examine changes in teacher and pupil perceptions over time.

Table 7: IPE methods overview

Research methods	Data collection methods	Participants/data sources	Data analysis methods	Research questions	Implementation/logic model relevance
------------------	-------------------------	---------------------------	-----------------------	--------------------	--------------------------------------

Evidence review	Literature search	Desk based research			
Discussion with developers	Phone interview	Programme developers	Thematic inductive analysis	RQb, c	Teacher Development Programme and Curriculum Assessment Programme
Observations of six training events	Non-participant observation	EM trainers, teachers	Review of fieldwork notes	RQb	Teacher-level outcomes; quality and impact of training
School visits	Observations of EM lesson	Teacher	Thematic inductive analysis	RQa–g	Teacher- and pupil-level outcomes
School visits	Interviews	English Mastery Lead and teacher(s)	Thematic inductive analysis	RQa–f, h	Teacher- and pupil-level outcomes
School visits	Focus groups	KS3 pupils	Thematic inductive analysis	RQg	Pupil-level outcomes (self-efficacy, motivation, engagement)
Pre- and post-intervention teacher survey	Online survey	Teachers	Descriptive statistics	RQa–f	Pre: business as usual, contextual factors; post: teacher- and pupil-level outcomes

Costs

Cost information was provided by the developer at the end of the trial to assess the typical cost to a school wanting to implement English Mastery. Costs per pupil per year were calculated over a three-year period based on EEF guidance published in February 2023. All figures are presented in the relevant section below.

Timeline

Table 8: Timeline

Dates	Activity	Staff responsible / leading
Dec 21–Jan 22	Set-up meetings, ethics approval	All
Feb 22	Design IPE instruments, IDEA workshop, evidence review	SHU
Feb–Jun 22	Data collection from schools (Cohort 1) Pre-intervention teacher survey (Cohort 1)	SHU/schools
Apr 22	Trial registration	SHU
Jun 22	Protocol published	SHU
Jun 22	Randomisation 1	SHU
Sep 22	Teacher training/observations (Cohort 1)	Ark/SHU
Sep 22	Pupil/class data collection Year 7 (Cohort 1)	SHU/schools
Oct 22–Jun 24	Delivery of English Mastery (Cohort 1)	Ark/schools
Jan 23–Mar 24	Conduct IPE school visits (Cohort 1)	SHU/schools
Feb–Jun 23	Data collection from schools (Cohort 2) Pre-intervention teacher survey (Cohort 2)	SHU/schools
Jun 23	Randomisation 2	SHU
Jun–Sep 23	Teacher training/observations (Cohort 2)	Ark/SHU
Sep–Dec 23	Pupil/class data collection Year 8 (Cohort 1) and Year 7 (Cohort 2)	SHU/schools
Oct 23–Jun 25	Delivery of English Mastery (Cohort 2)	Ark/schools
Jan 24–Mar 25	Conduct IPE school visits (Cohort 2)	SHU/schools
Jun 24	Outcome testing (Cohort 1) Post-intervention teacher survey (Cohort 1)	SHU/schools
Aug 24	Publish SAP	SHU/EEF
Feb 25	NPD application	SHU
Jun–Jul 25	Outcome testing (Cohort 2) Post-intervention teacher survey (Cohort 2) Collect fidelity data	SHU/schools
Sep–Dec 25	Data analysis	SHU
Jan 26	Report first draft	SHU

Impact evaluation results

Participant flow including losses and exclusions

Figure 2: Participant flow diagram

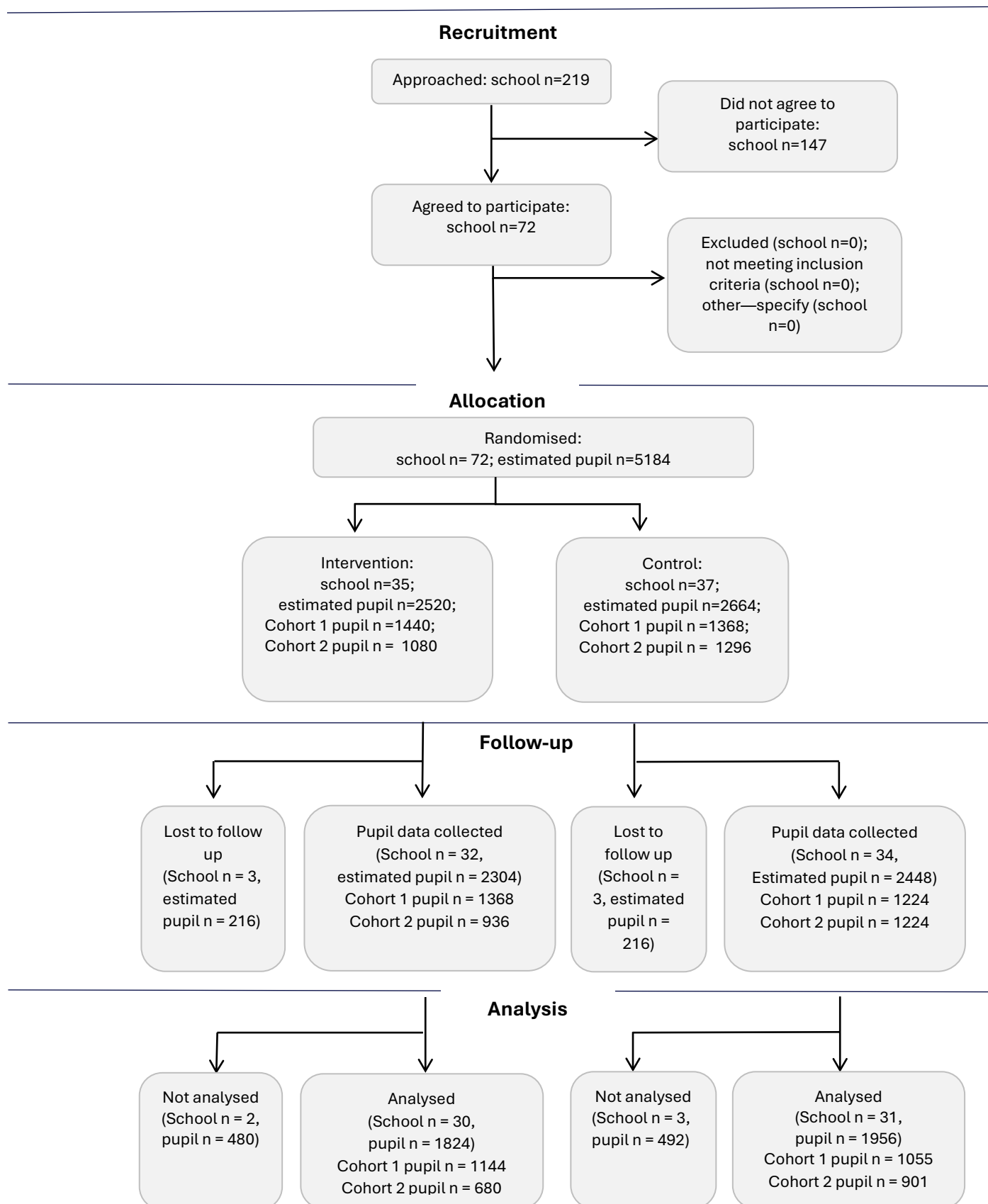


Table 9 shows the power calculations at the protocol, randomisation and analysis stages. The pre and post-test correlations refer to the baseline and endline primary outcomes (attainment in English) described above. As discussed, the original target was to recruit 100 schools, which is the projected sample size used in the protocol power calculations. The protocol MDES was 0.18 standard deviations for the overall sample and 0.19 for the FSM subsample. After adapting the trial design to recruit a second delivery cohort, the number of schools randomised was 72. Based on the same assumptions as the protocol power calculations, the lower sample size led to a less sensitive design powered to detect larger effect sizes than with the sample size assumed in the protocol and at randomisation (MDES overall = 0.21, MDES FSM = 0.22).

Table 9: Minimum detectable effect size at different stages

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.18	0.19	0.21	0.22	0.21	0.22
Pre-test/post-test correlations	Level 1 (pupil)	0.70	0.70	0.70	0.70	0.71	0.64
	Level 2 (class)	0.90	0.90	0.90	0.90	0.81	0.71
	Level 3 (school)	0.40	0.40	0.40	0.40	0.58	0.59
Intracluster correlations (ICCs)	Level 2 (class)	0.40	0.40	0.40	0.40	0.17	0.17
	Level 3 (school)	0.10	0.10	0.10	0.10	0.10	0.06
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		2	2	2	2	2	2
Average cluster size		12	3	12	3	9.4	3.2
Number of schools	Intervention	50	50	35	35	30	29
	Control	50	50	37	37	31	30
	Total:	100	100	72	72	61	59
Number of classes	Intervention	300	300	210	210	194	169
	Control	300	300	222	222	207	182
	Total:	600	600	432	432	401	351
Number of pupils	Intervention	3600	900	2520	630	1824	597
	Control	3600	900	2664	666	1956	515
	Total:	7200	1800	5184	1296	3780	1112

Of the 72 schools taking part, 62 completed outcome testing, one of which had no baseline data for its pupils, an unusual occurrence given the low missing data rates normally found in NPD data. The analysis sample therefore comprises 61 schools. The FSM subsample was 59 schools of which two had no pupils in the study sample classed as eligible for FSM in the NPD data. Despite this attrition, the MDES for the overall sample remained at 0.21 for the analysis sample. This is explained by the observed class-level ICC (0.17) being substantially lower than expected (0.40).

Attrition

For the trial overall, an estimated 27% of pupils—1,404 of 5,184—were not analysed. Attrition was slightly higher among schools in the intervention group (27.6%) compared to the control group (26.6%). One of the 62 schools that completed endline testing had no pupils with any baseline data and was therefore excluded from the analysis sample. As they were not analysed these pupils are counted within the attrition figure despite completing endline assessments.

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

		Intervention	Control	Total
Number of pupils	Randomised	2520	2664	5184
	Analysed	1824	1956	3780
Pupil attrition (from randomisation to analysis)	Number	696	708	1462
	Percentage	27.6%	26.6%	27.1%

Pupil and school characteristics

Table 11 profiles the schools in the randomisation sample. Recruitment was targeted in three regions where the Accelerator Fund was active but eligibility was later extended to other areas to increase the sample size. In the end, the majority of schools were recruited in the non-Accelerator Fund areas. This was the case for both the intervention (63%, $n = 22$) and control groups (62%, $n = 23$). Please note that the figures presented relate only to secondary schools. Middle schools are excluded here as it is not clear which ones would have been eligible for this study. Data is from 2021, the most recent available at the start of the trial, and was obtained from DfE via their school information website.

Around two-thirds of schools were rated as ‘good’ in their most recent OFSTED inspection, higher than the national average (58%) in the most recent data available at the time that the trial started (2021). ‘Outstanding’ schools were more likely to be in the control group ($n = 3$) than intervention ($n = 1$), and ‘Inadequate’ schools were more likely to be in the intervention group ($n = 5$, control $n = 1$). As a reminder, OFSTED rating was included as a stratifier in the minimisation process but this was treated as a binary variable, with outstanding and good schools grouped together into one category, and schools rated as requiring improvement or inadequate combined into one group. This was necessary given the low number of schools expected to have outstanding and inadequate ratings and led to this imbalance once the categories are disaggregated. Just over half of schools (51.4%) reported using some form of setting or streaming for English in KS3. National figures for school setting and streaming policy are not available so it is not possible to assess sample representativeness on this measure.

A higher proportion of schools in the control group (21.6%) were located in rural districts compared to the intervention group (8.6%). More intervention schools (82.9%) were academies or free schools than control schools (70%). The proportion of pupils classed as eligible for free school meals was slightly higher in intervention schools (23.3%) than in control schools (20.7%). These figures relate to the school overall as opposed to the pupils in the study sample.

Table 11: Baseline characteristics of groups as randomised

School-level (categorical)	National- level mean	Intervention group		Control group	
		n/N (missing)	Count (%)	n/N (missing)	Count (%)
East Mids./Humber	12%	3/35	3 (8.6%)	2/37	2 (5.4%)
West Mids.	13%	8/35	8 (22.9%)	9/37	9 (24.3%)
North	7%	2/35	2 (5.7%)	3/37	3 (8.1%)
Non-Accelerator Fund	68%	22/35	22 (62.9%)	23/37	23 (62.2%)
Outstanding	20%	1/35 (1)	1 (2.9%)	3/37	3 (8.1%)
Good	58%	23/35 (1)	23 (67.6%)	24/37	24 (64.9%)
Requiring improvement	15%	5/35 (1)	5 (14.7%)	9/37	9 (24.3%)
Inadequate	7%	5/35 (1)	5 (14.7%)	1/37	1 (2.7%)
Sets/streams	N/A	18/35	18 (51.4%)	19/37	19 (51.4%)
Rural	14.4%	3/35	3 (8.6%)	8/37	8 (21.6%)
Academies/free schools	36.9%	29/35	29 (82.9%)	26/37	26 (70.3%)
Local authority	46.9%	6/35	6 (17.1%)	11/35	11 (29.7%)
School-level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)
%FSM	24.7%	33/35 (2)	23.3 (11.4)	33/37 (4)	20.7 (10)

Table 12 below shows the baseline characteristics of the 61 schools in the analysis sample. For the school-level variables discussed above there is little change from the randomisation sample. Some imbalance exists in terms of school region, largely due to the small numbers recruited in certain areas. There is no evidence of an effect on the trial results. Imbalance on OFSTED ratings was already present in the randomisation sample and became more pronounced due to attrition. Four of the five intervention schools that did not complete endline testing were rated as 'good' compared to only two of the six control schools. Despite this, as OFSTED rating (in binary format) was used as a control in the analysis models along with prior attainment at the pupil, class and school levels, there is no evidence of this affecting the trial results. The percentage of FSM pupils in each school overall also remains very similar to that observed at randomisation, with more in the intervention group (22.3%) compared to the control group (19.6%). Both figures are below the national average for secondary schools (24.7%).

Prior attainment data is only available for the analysis sample as this was accessed from NPD at the end of the evaluation. Separate figures are reported for KS2 Reading and GPS, which are combined to create a baseline measure for the primary outcome in this study but assessed separately in statutory tests. For KS2 Reading, the effect size, calculated by dividing the

mean difference by the pooled standard deviation, is -0.02, while KS2 GPS scores were marginally higher among control pupils (effect size = 0.07). This slight imbalance is controlled for by including prior attainment in the analysis models. For the combined baseline measure, the effect size is -0.046 (0.09–0.17/1.83).

Table 12: Baseline characteristics of schools as analysed

School-level (categorical)	National- level mean	Intervention group		Control group		
		n/N (missing)	Count (%)	n/N (missing)	Count (%)	
East Mids./Humber	12%	2/30	2 (6.7%)	1/31	1 (3.2%)	
West Mids.	13%	8/30	8 (26.7%)	8/31	8/31 (25.8%)	
North	7%	1/30	1 (3.3%)	3/31	3 (9.7%)	
Non-Accelerator Fund	68%	19/30	19 (63.3%)	19/31	19 (61.3%)	
Outstanding	20%	1/30	1 (3.3%)	3/31	3 (9.7%)	
Good	58%	19/30	19 (63.3%)	22/31	22 (71%)	
Requiring improvement	15%	5/30	5 (16.7%)	5/31	5 (16.1%)	
Inadequate	7%	5/30	5 (16.7%)	1/31	1 (3.2%)	
Sets/streams	N/A	14/30	14 (46.7%)	16/31	16 (51.6%)	
Rural	14.4%	3/30	3 (10%)	6/31	6 (19.4%)	
Academies/free schools	36.9%	25/30	25 (83.3%)	22/31	22 (71%)	
Local authority	46.9%	5/30	5 (16.7%)	9/31	9 (29%)	
School-level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	
%FSM	24.7%	29/30 (1)	22.3 (10.7)	29/31 (2)	19.6 (9.1)	
Pupil-level (continuous)*		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size
KS2 Reading		1824	31.1 (10.22)	1956	31.3 (10.14)	-0.02
KS2 GPS		1824	42.3 (14.52)	1956	43.3 (14.67)	-0.07
Combined baseline measure		1824	0.09 (1.83)	1956	0.17 (1.83)	-0.046

Outcomes and analysis

Primary analysis

All results shown in this section are from the headline model containing pupil-, class-, and school-level baseline covariates and the three variables used in the stratification—OFSTED rating, region, and school setting/streaming policy—unless specified otherwise. Table 13 displays the results from analysis of the primary outcome with all pupils in the ITT sample who completed the endline assessment. The primary outcome measure was the raw score for GL PTE13, and the baseline measure was the sum of the standardised scores from the KS2 Reading and GPS tests. Histograms for both of these measures can be found in Appendix E (Additional Appendices). The distribution of both scores for the primary outcome is similar with both samples having a positive skew. Mean test scores for pupils in the control group (27.24) were marginally higher than for the intervention group (26.40), and the point estimate of the effect size (-0.04, CIs -0.169, 0.083) is negative but very small, equivalent to zero months of additional progress for the intervention group. The trial was not powered to detect an effect of this magnitude. Consistent with this, the observed p-value (0.502) does not provide evidence of an effect. Furthermore, the confidence intervals range from two months less progress to one month of additional progress, illustrating the uncertainty around the point estimate.

Table 13: Analysis of primary outcome for ITT sample

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
GL PTE13	1824 (501)	26.4 (25.765; 27.044)	1956 (651)	27.24 (26.637; 27.844)	3780 (1824; 1956)	-0.04 (-0.169; 0.083)	0.502

Secondary analysis

Table 14 shows the results from analysis of the two secondary outcomes with all pupils in the ITT sample who completed the endline assessment. The two measures are the raw scores for the Reading and GPS components of the GL PTE13. Histograms for these measures are presented in Appendix E (Additional Appendices). Both secondary outcome measures have a negatively skewed distribution in both the intervention and control samples. The corresponding KS2 tests were used as baseline measures. Both secondary outcomes have very small negative effect sizes (Reading: -0.05; GPS: -0.03) representing one month of negative progress and zero months of additional progress respectively. The wide confidence intervals (Reading: CIs = -0.188; 0.091; GPS: CIs = -0.143; 0.091) and high p values (Reading: p = 0.493; GPS: p = 0.660) also attest to the trial design being unable to detect such small effect sizes with certainty.

Table 14: Analysis of secondary outcomes for ITT sample

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
GL PTE13 Reading	1826 (499)	12.3 (11.94; 12.659)	1960 (647)	12.68 (12.343; 13.018)	3786 (1826; 1960)	-0.05 (-0.188; 0.091)	0.493
GL PTE13 GPS	1834 (491)	6.83 (6.666; 6.996)	1961 (646)	6.86 (6.703; 7.018)	3795 (1834; 1961)	-0.03 (-0.143; 0.091)	0.660

Analysis in the presence of non-compliance

The developer provided data on the seven indicators which were combined into a single compliance measure as described above. Of the 30 intervention schools in the analysis sample, ten were deemed compliant with the intervention, with 648 pupils in the analysis sample. The effects on non-compliance were analysed through a 2SLS model. Table 15 summarises the results of the first stage model, with compliance to the intervention as the outcome. There is no discernible association between pupil baseline score and compliance and for the class baseline score the coefficient is negligible (0.01, CIs -0.003, 0.005, $p = 0.014$), yet for school baseline scores the coefficient is larger (-0.18, CIs -0.311, -0.044, $p = 0.009$), suggesting that schools with lower baseline scores were more likely to achieve compliance.

In the second stage of the 2SLS model (Table 16), the negative coefficient (-2.07, CIs -743, 3.29) for the compliance indicator shows that pupils in compliant schools were more likely to record lower scores on the outcome assessment compared to others in the study sample, including those in the control group and those in the intervention schools deemed not compliant. Converting this into an effect size by dividing by the unconditional standard deviation (13.98) produces an effect size of -0.15 (see Table 17), far larger than for the headline ITT model reported above (-0.04). However, the p -value (0.449) is high suggesting that this may be due to chance and the coefficient value (two marks on the test, which is scored on a scale from zero to 65) suggests that any differences are small. This is corroborated by the weak correlation between the primary outcome measure and the compliance variable (0.02).

Table 15: Results from 2SLS model of compliance—first stage

	Coefficient	SE	p-value	Lower 95% CI	Upper 95% CI
Pupil Baseline	0.00	0.002	0.604	-0.003	0.005
Class Baseline	0.01	0.003	0.014	0.002	0.013
School Baseline	-0.18	0.068	0.009	-0.311	-0.044
RSC region (ref other)					
East Midlands	0.32	0.195	0.101	-0.062	0.700
North	-0.02	0.069	0.737	-0.159	0.113
West Midlands	0.12	0.096	0.221	-0.071	0.305
Mixed ability	0.19	0.082	0.022	0.027	0.347
OFSTED binary	0.15	0.100	0.121	-0.041	0.350
Treatment Allocation	0.34	0.086	0	0.176	0.512
Constant	-0.40	0.202	0.05	-0.792	0.000

N = 3780; F-ratio = 3.12.

Table 16: Results from 2SLS model of compliance—second stage

	Coefficient	SE	p-value	Lower 95% CI	Upper 95% CI
Compliance	-2.07	2.737	0.449	-7.434	3.294
Pupil Baseline	4.98	0.160	0.000	4.669	5.296
Class Baseline	5.85	0.256	0.000	5.352	6.356
School Baseline	4.24	1.050	0.000	2.181	6.297
RSC region (ref other)					
East Midlands	3.16	2.428	0.193	-1.601	7.915
North	-3.51	0.950	0.000	-5.368	-1.644
West Midlands	-0.96	1.141	0.402	-3.192	1.281
Mixed ability	2.08	1.060	0.049	0.006	4.161
OFSTED binary	2.19	1.253	0.081	-0.269	4.641
Constant	21.88	2.479	0.000	17.026	26.743

Table 17: Effect size based on compliance coefficient in 2SLS model

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
GL PTE13	1824 (501)	26.4 (25.765; 27.044)	1956 (651)	27.24 (26.637; 27.844)	3780 (1824; 1956)	-0.15 (-0.531; 0.235)	0.449

Missing data analysis

Pupil attrition in this trial was 27%. Of the estimated 5,184 pupils in the randomisation sample, 3,780 completed the outcome assessment at the end of the evaluation period. Patterns in missing outcome data were examined through a logistic regression model. The outcome variable was a binary indicator where '1' represented missing outcome data and '0' represented outcome data not missing. Predictor variables from the headline ITT model were included to check whether any were significantly related to missing outcome data. The results are presented in Table 18. The odds ratios for baseline scores at both pupil (0.92, $p < 0.001$) and class (0.69, $p < 0.001$) level show that higher scores on the baseline test are associated with a lower likelihood of missing outcome data. In other words, pupils with lower KS2 English scores were more likely to have missing outcome data, as were pupils in classes with lower average KS2 English scores. This suggests that attrition led to an analysis sample of pupils with higher prior attainment than the ITT sample.

Multiple imputation was then carried out using the `mi impute` command in Stata. Data from the variables included in the headline analysis models was used to estimate outcome test scores where this was missing. Ten imputations were carried out and in each case the imputed outcome variable was used in place of the primary outcome measure to repeat the analysis on the overall sample. The sample size for these models increased to 4,620 and the point estimates ranged from -0.02 to -0.05. As the headline ITT effect size was -0.04, the similarity of these point estimates suggests that missing data did not substantially alter the trial results. The full results can be found in Appendix G (Additional Appendices).

Table 18: Missing data analysis

	Odds ratio	SE	p-value	Lower 95% CI	Upper 95% CI
Pupil Baseline	0.92	0.03	0.00	0.87	0.97
Class Baseline	0.69	0.02	0.00	0.64	0.74
School Baseline	0.65	0.20	0.16	0.36	1.18
RSC region (ref other)					
East Midlands	0.53	0.39	0.40	0.13	2.27
North	0.83	0.52	0.76	0.24	2.83
West Midlands	0.49	0.18	0.05	0.24	0.99
Mixed ability	1.28	0.39	0.41	0.71	2.32
OFSTED binary	1.06	0.38	0.86	0.53	2.14
Constant	0.16	0.11	0.01	0.04	0.61

Subgroup analyses

Subgroup analysis was conducted on pupils eligible for FSM as expected for all EEF trials. Restricting the sample to FSM pupils and repeating the analysis of the primary outcome produces a point estimate of -0.11 standard deviations (CIs: -0.259; 0.036, $p = 0.138$). This represents two months of negative progress for socioeconomically disadvantaged pupils in the intervention group. The trial was not powered to detect an effect of this size given the smaller subsample ($n = 1,112$), yet it is estimated that there was lower attrition among FSM pupils (14.4%) compared to the overall sample (27.1%). The range of possible effects also includes no additional progress, suggesting a degree of uncertainty about the result.

Analysis of this subgroup also included a model with the full sample and an interaction effect between FSM and intervention group membership. This model produced a smaller negative effect to that found from the analysis restricting the sample to FSM pupils (-0.05, CIs -0.141; 0.046, $p = 0.315$), more similar to the ITT point estimate (-0.04) with confidence intervals again straddling zero. Given the lack of evidence from the IPE about why the results for socioeconomically disadvantaged pupils may differ from those of other pupils, it is possible that the variation in these findings is due to the small size of the restricted FSM subsample.

Further analysis was performed on the EAL subgroup. The point estimate here was again negative (-0.08) with wide confidence intervals (-0.266; 0.107, $p = 0.402$). For pupils defined as having a Special Educational Need, the effect was also negative but smaller (-0.04, CIs: -0.224; 0.138, $p = 0.644$). The final subgroup analysed was pupils with lower prior attainment, defined here as the lowest quartile on the baseline measure for the primary outcome (KS2 English Reading and GPS combined). The effect size was again negative ($g = -0.07$, CIs: -0.264; 0.116, $p = 0.447$), so there was no evidence of any additional progress for pupils in intervention schools across any of the subgroups. These findings should be interpreted with the caveat that the trial was not powered for such small sample sizes so there is uncertainty about the precision of the estimates. The results from all subgroup analyses are presented in Table 19.

Table 19: Subgroup analysis of primary outcome

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
FSM	597 (203)	22.14 (21.083; 23.202)	515 (221)	23.38 (22.219; 24.538)	1112 (597; 515)	-0.11 (-0.259; 0.036)	0.138
FSM*Treatment	597 (203)	22.14 (21.083; 23.202)	3183 (784)	27.72 (27.241; 28.194)	3780 (597; 3183)	-0.05 (-0.141; -0.046)	0.315
EAL	341 (70)	25.7 (24.251; 27.145)	515 (154)	27.37 (26.18; 28.554)	856 (341; 515)	-0.08 (-0.266; 0.107)	0.402
SEN	344 (155)	18.19 (16.902; 19.475)	372 (187)	19.22 (17.98; 20.461)	716 (344; 372)	-0.04 (-0.224; 0.138)	0.644
Low prior attainment	423 (144)	13.86 (13.172; 14.544)	426 (1073)	13.83 (13.126; 14.536)	849 (423; 426)	-0.07 (-0.264; 0.116)	0.447

Additional analyses and robustness checks

Schools were recruited to this trial in two cohorts as described above. It was agreed that each should be examined separately as a sensitivity analysis to check for any differences in results. Table 20 shows that the point estimates for the two cohorts were noticeably divergent. For Cohort 1, membership of the intervention group is associated with a negative effect of -0.14 standard deviations (CIs: -0.279; 0.001, $p = 0.048$), whereas for Cohort 2 there is a small positive effect ($g = 0.04$, CIs: -0.186; 0.272, $p = 0.714$). Despite the clear difference between the results for the two cohorts, the effect size for Cohort 2 was smaller than the trial was powered to detect. One possible reason for this difference is changes to the delivery during the evaluation period. Intervention schools in the second cohort received a slightly updated version of the programme, as noted in the compliance section of this report. Surveys and interviews with teachers and Mastery Leads suggested that the programme had undergone considerable development, as highlighted in the introductory section, prior to and during the trial period. However, there is no specific evidence as to why this might have affected impact on pupil attainment.

Table 20: Analysis of primary outcome by delivery cohort

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Cohort 1	1144 (246)	25.9 (25.07; 26.731)	1055 (225)	27.72 (26.88; 28.555)	2199 (1144; 1055)	-0.14 (-0.279; -0.001)	0.048
Cohort 2	680 (255)	27.25 (26.259; 28.247)	901 (426)	26.68 (25.811; 27.552)	1581 (680; 901)	0.04 (-0.186; 0.272)	0.714

Estimation of effect sizes

The headline effect size of the ITT model (-0.04) was calculated by dividing the allocation coefficient, which is the mean difference between the intervention and control group outcome scores controlling for all covariates (-0.86), by the unconditional standard deviation (13.98) derived from the total variance from the outcome only model (195.4). The formula was presented above (Equation 5).

Estimation of ICC

The school-level ICC for the outcome-only model that provided the unconditional variance for the primary outcome analysis with the main ITT sample was 0.36 and the class-level ICC was 0.05. For the full model including all covariates, used to generate the coefficient for the allocation variable that provides the numerator for the effect size equation, the school-level ICC was 0.10, and the class-level ICC was 0.17. The observed class-level ICC was substantially lower than expected at protocol (0.40), although the school-level ICC was the same as the protocol assumption (0.10), as noted above. The formulae were presented above (Equations 6 and 7).

Table 21: Variance and ICC figures from analysis of primary outcome

Model	School variance	Class variance	Pupil variance	Total variance	School ICC	Class ICC
Outcome only	10.70	60.49	124.22	195.41	0.05	0.36
With baseline	11.30	7.13	76.18	94.61	0.12	0.19
Full model	8.99	7.09	76.18	92.26	0.10	0.17

Implementation and process evaluation results

This section reports findings from the IPE and is structured thematically according to the research questions. A full description of the methods and data sources used can be found in the IPE Methods section.

Changes to the intervention since the previous English Mastery trial

Surveys and interviews with teachers and Mastery Leads suggested that the programme had undergone considerable development, as highlighted in the introductory section, prior to and during the trial period. Teachers valued the developer's guidance on adapting Mastery materials to suit their pupils, including an emphasis on dual language texts for EAL pupils. It is clear from interviews with teachers that the delivery team was very responsive to school requests for guidance and suggestions for improvement of the programme. This was corroborated by an interview with the delivery team after the first year of the programme. During interviews conducted on school visits, respondents reported having received significant support with respect to assessments and co-planning and interview data also suggested that Ark had expanded the guidance on adapting lessons to meet the needs of a range of pupils. There were mixed reviews of how far the programme included contemporary texts, some expressing that the range of texts was rich and worked well for pupils, others suggesting that this could be developed further. Overall, the Literary Heritage strand was praised for its depth and challenge, to which pupils responded to well.

a) Are any changes in self-reported teacher workload associated with English Mastery?

The teacher surveys administered at baseline and the end of the first and second years of the delivery period (end of Year 7 and end of Year 8 for participating pupils) included a set of items on teacher workload. These questions asked respondents to specify how long they spent each week on certain activities. An overall comparison across control and intervention schools shows that at baseline teachers in intervention schools spent an average of 35.55 hours on teaching English, whereas teachers in control schools spent 32.77 hours (see Table 22). For intervention schools, this reduced to 33.85 hours at the end of Year 7 and 31.85 hours at the end of Year 8, in contrast to 28.18 hours (end Year 7) and 29.32 hours (end Year 8) for control schools. This suggests that by the end of the trial, there was a slightly larger reduction in workload for intervention schools (3.7 hours) than control schools (3.45 hours). As the evaluation did not include any fieldwork in control schools, there is no evidence on why these respondents reported lower hours teaching English at baseline or at mid- and post-intervention. It is worth noting that while intervention schools reported a continued decrease in time spent teaching English over the observation period, for control schools the pattern was not consistent.

From teacher interviews in intervention schools, self-reported changes in workload associated with English Mastery were mixed. Most respondents felt that the English Mastery materials allowed them to focus more on lesson delivery, co-planning, and adaptation rather than preparing materials. This was seen as improving the quality of teaching overall, yet while teachers acknowledged that less time was spent preparing materials, the materials often required adaptation to make implementation successful. It should be noted here, however, that the perceived positive impact on the quality of teaching did not lead to any significant impact based on the impact evaluation.

'Having these brilliant resources does reduce your workload, but at the same time it doesn't. If you want to do anything well, you are going to have to dedicate a little bit of time and effort to it. So, because the quality of the lessons has improved, the quality of teaching has improved, the quality of resource has improved—we could have carried on teaching the same lessons and it would have taken us less time to do that, but the quality of the teaching and the lessons wouldn't have been improving' (Mastery Lead, Year 7).

'The teacher has been able to focus on the co-planning, the preparation aspects, the implementation, rather than having to stay up all night writing schemes of work that probably would not be as good' (Mastery Lead, Year 7).

Benefits were seen for trainee teachers or those who are less experienced teachers. This finding was voiced by five of ten Mastery Leads interviewed during school visits:

'I have a trainee teacher who is trying to develop her subject knowledge on multiple texts that she has not taught before, and there is so much just on demand that she can flick through what's relevant to everything that we're teaching in every lesson. So it's there and it's not a wasteful internet search of trying to find different things. There is something on every lesson, or every piece of Mastery content that there is. It saves a lot of time' (Mastery Lead, Year 8).

During school visits, six English Mastery Leads explained how familiarisation with English Mastery involved an investment of time but once it was established it saved time for teachers as their knowledge of the programme increased. Teachers often saw this as an investment of time that would improve their teaching. For example, where norm-referenced assessments were not usual practice in the school, this took time to put into practice. One teacher explained:

'We feel confident in the sense that the system appears to be quite robust and thorough, and it doesn't mean that it's not a challenge to engage with it, because it is quite challenging using standardisation materials and a very detailed mark scheme, but I would say that's a good challenge' (Teacher, Year 8).

Survey data reflected similar trends in workload across intervention schools (see Table 22) where a substantial reduction in teacher time spent preparing lesson plans was noted, from 3.67 hours at baseline to 2.16 hours at the end of Year 7 and 2.57 at the end of Year 8. In contrast, the time spent preparing lessons plans for control schools decreased from 2.81 at baseline to 2.08 at the end of Year 7, before increasing again to 3.71 at the end of Year 8. Intervention schools reported 2.57 hours per week at baseline spent developing resources, 1.52 hours per week at the end of Year 7, and 1.54 hours per week at end of Year 8. The control group also reported spending less time on this, from 2.66 hours per week at baseline to 1.86 at the end of the evaluation period. Interestingly, despite some initial concerns about the time taken to adjust to the English Mastery approach expressed by staff during school visits, time spent by intervention teachers on administering and marking work fell from 2.51 hours per week at baseline to 2.22 hours per week at the end of Year 8. The control group reported an increase in time spent: from 2.47 hours per week at baseline to 2.91 at the end of Year 8. Co-planning is a crucial feature of the English Mastery programme. For intervention schools, time spent co-planning increased slightly from baseline (1.69 hours) to the end of Year 7 (1.71 hours), before decreasing slightly at the end of Year 8 (1.56 hours). In contrast, less time was spent co-planning in control schools, dropping from 1.84 hours at baseline to 1.01 hours at the end of Year 8.

Table 22: Findings from teacher workload questions analysed at the school level

	Group*	Baseline	End of Year 7	End of Year 8
Overall teaching	Intervention	35.55	33.85	31.85
	Control	32.77	28.18	29.32
Teaching English	Intervention	5.74	5.87	4.37
	Control	5.06	4.82	4.89
Preparing lesson plans	Intervention	3.67	2.16	2.57
	Control	2.81	2.08	3.71
Developing resources	Intervention	2.57	1.52	1.54
	Control	2.66	2.13	1.86
Administering and marking work	Intervention	2.51	2.45	2.22
	Control	2.47	2.76	2.91
Non-curriculum administration	Intervention	2.45	1.42	2.12
	Control	2.27	1.64	1.82
Co-planning	Intervention	1.69	1.71	1.56
	Control	1.84	1.34	1.01
External co-ordination	Intervention	1.46	0.83	1.3
	Control	2.12	1.37	0.69
Overall workload change	Intervention	2.55	2.01	2.2
	Control	2.37	2.35	2.38

* Sample of 13 schools for the intervention group, 24 schools for the control group. Source: Evaluation teacher survey.

b) To what extent does the English Mastery training, its scheduling, and materials equip teachers and leaders to implement a coherent and cumulative English curriculum?

Alongside the schemes of work provided, the English Mastery programme resources are accessible to participating teachers through the MyMastery platform. My Mastery resources are based on the four pillars of Ark Curriculum Plus: accumulation of knowledge, discrete grammar teaching, tier two vocabulary, and the use of standardised, norm-referenced assessments. Interviews with Mastery Leads and teachers suggested that these pillars supported the delivery of a coherent and cumulative English curriculum, improvements in teacher subject knowledge, and the opportunity to engage in co-planning. Where co-planning had been scheduled regularly, it was seen as a positive move away from more individual ways of working, raising the quality of teaching through detailed assessment and teacher collaboration. However, time for co-planning was limited in some schools due to staff time constraints. The programme was seen by interviewees as improving teacher subject knowledge, particularly in terms of teaching grammar and building on what pupils had learned at KS2, which had been identified as a 'weakness' in their English curriculum by a number of schools.

'I don't think that we had as much knowledge as we needed to be able to do a good job of it on our own' (Mastery Lead, Year 7).

'And because a lot of us don't have the knowledge of KS2, it was also at what point do you go in?' (Mastery Lead, Year 7).

Co planning with English Mastery resources was a feature of the programme that schools valued. The collaboration that took place during co-planning meetings was seen as beneficial and commented on positively by English Mastery Leads.

‘The whole thing is that it really encourages collaborative planning and collaborative working’ (Mastery Lead, Year 8).

Support from senior management in facilitating time for co-planning was essential for its success. One Mastery Lead saw the potential value in co-planning but did not have time to schedule it. A significant number of schools, however, saw it as a significant change in practice that led to improvements in teaching and assessment in English:

‘We’ve got really good support from senior leaders with implementing co-planning, especially, so we’ve all got that protected time. It’s all on our timetable, it’s not going to be used [for anything else], and we all join together every fortnight to look ahead at things like progress checks, assessments, or even just smaller lessons that are coming up in the future’ (Mastery Lead, Year 7).

‘I think even a simple thing like the co-planning, that is not something we would have done before. I don’t think that I’ve ever had a meeting and sat down where we actually talk about the misconceptions or ideas that are coming up, or any adaptations that we might want to take together. I think that it’s making us quality check things a lot more in terms of what we want to do with the resources and why’ (Mastery Lead, Year 7).

The support from the delivery team was seen as a key feature of the programme that was supportive and tailored to the schools. Table 23 below provides further insights into which parts of the English Mastery training were seen as most beneficial to teachers. Elements that were highly rated were:

- induction training—with a mean score of 4.14 at the end of Year 7 and 4.06 at the end of Year 8;
- school visit support—with a mean score of 4.1 at the end of Year 7 and at the end of Year 8; and
- My Mastery teaching overviews—with a mean score of 4.02 at the end of Year 7 and 3.96 at the end of Year 8.

That said, there were some elements of the programme where scores decreased slightly. The mean score for ‘Scheduling’ decreased slightly from 3.89 at the end of Year 7 to 3.78 at the end of Year 8, and the mean score for remote development sessions decreased from 3.95 at the end of Year 7 to 3.67 at the end of Year 8, however these were small changes with scores still relatively high overall. The items ‘Materials’ and ‘My Mastery Teaching CPD Modules’ remained constant throughout both years. Overall, there was a slight reduction in satisfaction across most items at the end of Year 8 compared to Year 7, although responses remained broadly positive.

School visit support was described favourably by all Mastery Leads during fieldwork interviews and this seemed to be a major strength of the programme. One Mastery Lead told us:

‘So Ark have supported us very well. We had a visit from the subject development lead and [they were] very helpful, very complimentary but also very helpful in terms of giving us directions for how to improve things, how to ensure assessments across department and ‘is it fit for purpose?’ and things like that. [They were] very, very supportive, but also not just ‘you’re doing a great job, keep it up!’, you know, [they] very much had the big picture in mind in terms of how we can improve.’

Ark staff were seen as very knowledgeable and interviewees recounted how they listened to any issues that schools were experiencing, offering support and advice on aspects of the programme such as marking and assessment. Ark staff were also seen as acting quickly on any feedback or suggestions from schools. One Mastery Lead described the school visits as elevating the Mastery programme to a ‘whole-school development tool’. The holistic approach to support provided by Ark, individualised to the school, was seen as one of the most beneficial aspects.

Table 23: Teacher perceptions of English Mastery training

	End of Year 7			End of Year 8		
	N	Mean	SD	N	Mean	SD
Overall	20	3.97	0.67	20	3.89	0.63
Materials	22	3.94	0.56	20	3.93	0.50
Scheduling	21	3.89	0.46	19	3.78	0.58
Induction training	22	4.14	0.69	20	4.06	0.82
Remote development sessions	19	3.95	0.66	18	3.67	0.75
School visit support	22	4.1	0.65	19	4.1	0.82
MyMastery—teaching overviews	22	4.02	0.5	18	3.96	0.71
MyMastery—teaching mastery CPD modules	21	3.9	0.56	18	3.81	0.64

Source: Evaluation teacher survey. Responses: 1, 'terrible'; 2, 'poor'; 3, 'fair'; 4, 'good'; 5, 'excellent'.

- c) In what ways has English Mastery led to changes in practice in the teaching of English, including for supporting the needs of different pupil groups (such as EAL, SEND and FSM, and those on traditional/foundation pathways)?

Changes in practice in the teaching of English

In considering any changes in practice in the teaching of English, it is worth understanding schools' motivations for joining the trial. When English Mastery Leads were asked this question, schools wanted to be confident that the English curriculum had the required consistency, progression, and coverage and provided strong preparation for GCSE English:

'I think that it meets all of the requirements of the National Curriculum and what we really liked about it was the three different strands it offers. So the Literary Heritage, the Writing Mastery and also the Reading for Pleasure. It kind of ticks every box in terms of the genres that they produce for every year group' (Mastery Lead).

'We joined it because we wanted a fully resourced curriculum that would take the pressure off teachers, and also to make sure that students are getting a rich deal, if you like, to make sure that they are learning a rich literary heritage and also ensuring that accuracy as well' (Mastery Lead).

For a number of schools, the explicit teaching of grammar and punctuation, its accurate application, and teacher subject knowledge were seen as weaknesses that would be addressed through English Mastery. These motivations for joining the trial recur in both survey and interview data. Where the programme was perceived by teachers as implemented successfully in their school, it was seen as leading to overall improvements in the quality of teaching. This was particularly noted where English Mastery Leads were working with less experienced teachers or those who were dual subject specialists:

'The less experienced teachers—it's made them stronger teachers' (Mastery Lead).

The teacher survey measured teacher confidence on a five-point scale ranging from no confidence ('1') to high confidence ('5'). In areas such as formative assessment and providing effective feedback there appears to be little impact on teacher self-confidence. During interviews on school visits, feedback was seen as an effective strategy embedded in the programme that supported progress and confidence of pupils. However, some teachers reported already feeling confident in this.

In terms of other areas important to English Mastery, the mean score for use of summative assessment to inform planning for intervention schools was 4.05 at baseline. This increased to 4.26 at the end of Year 7 and 4.34 at the end of Year 8,

reflecting increased confidence. However, for control schools, the mean score (4.39) was slightly higher than intervention schools at baseline. This increased slightly to 4.38 at the end of Year 7 and 4.48 at the end of Year 8 suggesting that confidence increased across both groups. Likewise, confidence in providing effective feedback rose from 4.17 at baseline to 4.28 at the end of Year 7 and 4.24 at the end of Year 8 for intervention schools. While this was a smaller increase, these figures show fairly high levels of teacher confidence. In comparison, confidence in providing effective feedback among control schools remained relatively stable, increasing from 4.3 at baseline to 4.38 at the end of Year 7 before decreasing slightly to 4.33 at the end of Year 8.

Teachers in intervention schools judged that their subject knowledge overall had improved over the two years, rising from 4.23 at baseline to 4.49 at the end of Year 7 and 4.52 by the end of Year 8. Control schools, however—from a higher baseline of 4.48—moved to 4.51 by the end of Year 7 to 4.63 by the end of Year 8. In the question to teachers regarding improvements to their ‘English pedagogy overall’, where the starting point was 4.1 in the pre-intervention survey, intervention teachers then moved to 4.28 by the end of Year 7 and 4.41 by the end of Year 8. A higher score at pre-intervention for control schools indicated 4.13, which moved to 4.18 by the end of the first year of the intervention to 4.46 by the end of the second year.

Greater improvements in confidence in the teaching of English for intervention schools were found in teaching grammatical concepts, teaching vocabulary, motivating pupils to enjoy English, and for teaching different groups of pupils. These reflect the pillars of the Ark English Mastery approach. In teaching grammatical concepts, from a baseline of 3.76, teachers moved to 4.11 by the end of Year 7 to 4.28 by the end of Year 8. In comparison, for control schools at the end of the second year the mean was 4.04. In teaching vocabulary, a baseline mean of 4.09 rose to 4.33 by the end of Year 7 and 4.64 by the end of Year 8. In contrast, teachers in control schools at the end of Year 8 scored 4.52 for confidence in teaching vocabulary. Increased confidence to motivate pupils to enjoy English was noted: at baseline this was 4.04 in intervention schools, rising to 4.21 at the end of Year 7 and to 4.64 by the end of Year 8. Control schools gave a score of 4.18 for this aspect of English teaching. What is notable is the steady incremental gains in confidence across the two years of the intervention for these aspects of English teaching.

Table 24: With respect to teaching English, how confident are you in the following?

	Group*	Baseline	End of Year 7	End of Year 8
Using formative assessment to inform planning	Intervention	4.27	4.36	4.30
	Control	4.42	4.39	4.53
Using summative assessment to inform planning	Intervention	4.05	4.26	4.34
	Control	4.39	4.38	4.48
Providing effective feedback to students	Intervention	4.17	4.28	4.24
	Control	4.30	4.38	4.33
Your overall English subject knowledge	Intervention	4.23	4.49	4.52
	Control	4.48	4.51	4.63
Your overall English pedagogy	Intervention	4.10	4.28	4.41
	Control	4.13	4.18	4.46
Teaching of canonical texts	Intervention	3.81	4.10	4.37
	Control	4.00	4.06	4.36
Teaching grammatical concepts	Intervention	3.76	4.11	4.28
	Control	3.82	3.88	4.04
Teaching vocabulary	Intervention	4.09	4.21	4.64
	Control	4.19	4.27	4.52
Motivating pupils to enjoy English	Intervention	4.04	4.33	4.46
	Control	4.18	4.24	4.18

* Sample of 13 schools for intervention group; sample of 24 schools for control group. Source: Evaluation Teacher Survey. Responses: 1, 'no confidence' to 5, 'very confident'.

Changes in practice in the teaching of English including FSM, EAL, and pupils with SEND

The teacher survey also asked respondents about the impact of the English Mastery programme on their confidence in teaching pupils identified as FSM, EAL, or SEND. During the course of programme delivery, Ark developed training and teaching materials to support teachers in tailoring the resources to pupil needs. Teachers were also advised to adapt materials accordingly to suit their classes. Commenting on the foundation pathway, one Mastery Lead said:

'I don't think that supplying just a foundation pathway is going to tick every box in terms of meeting those needs. I think that comes to us in how we adapt that, which I think—through observations and what we're doing together—I think that we can see that it is working' (Mastery Lead).

Table 25 shows changes in confidence in aspects of teaching in both intervention and control schools. Responses were again recorded on a scale from '1' (no confidence) to '5' (very confident). Differentiating learning activities and the 'chunking' of content in English lessons was seen by some as an improvement for pupils who may experience difficulty:

'I've been on a learning walk and to see both sets of lessons, and I can see how both sets of lessons are enabling students to learn the same content but in a different way. So, I've appreciated that level of differentiation' (Mastery Lead).

Teachers reported little difference over time in their confidence in teaching pupils eligible for FSM, with both intervention and control schools having relatively high baseline scores (4.14 and 4.15 respectively) increasing slightly to 4.31 and 4.3 by the end of Year 8. This suggests that confidence in supporting FSM pupils was already well established and remained broadly stable throughout the trial.

In contrast, more noticeable changes were observed in relation to teaching pupils with EAL. For intervention schools, mean scores increased from 3.15 at baseline to 3.47 at the end of Year 7, and further to 3.57 by the end of Year 8. While control schools also showed some improvement over time, the increase in mean scores for these schools was smaller, rising from 3.17 at baseline to 3.37 by the end of Year 8.

A similar pattern emerged for teaching pupils with SEN. Although baseline scores were comparable across intervention and control schools (3.83 and 3.84 respectively), teachers in the intervention schools reported a more substantial increase by the end of Year 8, reaching a mean score of 4.17, compared to 3.96 in control schools.

Table 25: With respect to teaching English, how confident are you in the following?

	Group	Baseline	End of Year 7	End of Year 8
Teaching FSM pupils	Intervention	4.14	4.19	4.31
	Control	4.15	4.21	4.3
Teaching EAL pupils	Intervention	3.15	3.47	3.57
	Control	3.17	3.22	3.37
Teaching SEN pupils	Intervention	3.83	3.86	4.17
	Control	3.84	3.84	3.96

* Sample of 13 intervention schools and 24 control schools. Source: Evaluation Teacher Survey. Responses: 1, 'no confidence' to 5, 'very confident'.

Teacher interviews during school visits suggested that the Literary Heritage and Writing Mastery units were seen as beneficial for EAL pupils. Over the two years of the evaluation, the developer increased the emphasis on oracy and adaptable materials. This was reflected in teacher interviews as particularly helpful for EAL pupils:

'I think like for Literary Heritage, I think any focus on story makes it really accessible for EAL and I think the Literary Heritage units do that really well and I think the Writing Mastery builds their grammar skills' (Mastery Lead).

'I think that's where the oracy strands are coming in. That's really helping with that. I think with oracy and vocabulary, it's very much exposure and not necessarily about their attainment at that point. So I know Ark are developing each unit started from Year 7 to try and put more speaking practice into every unit' (Mastery Lead).

The emphasis on vocabulary was seen as positive. One school praised the tool that allowed them to translate PowerPoint slides into home languages.

'The focus on vocabulary within the curriculum both in the Writing Mastery and in the Literary Heritage and in the Reading for Pleasure—I think that has been really valuable for EAL students' (Mastery Lead, Year 8).

'The dual coding definitely helps them access the powerful knowledge in terms of the subject-specific vocabulary' (Mastery Lead, Year 8).

When commenting on the impact of the programme and their confidence in teaching pupils with SEND in Year 7, the structure of English Mastery sessions was well received in that lessons are broken down into small steps, allowing pupils and teachers to see progress and areas of difficulty:

'So we are finding that lower set that they are all SEN students in there who have always struggled with reading, so we've been able to give them the right teaching, rather than a one size fits all' (Mastery Lead).

'The programme definitely allowed us to take snapshots of that moment in time when students need our support and how we can develop those students that are doing well with it, how we can carry on supporting them to move on, but also stop and re-address it for those students that need that little bit more' (Mastery Lead).

d) How have the English Mastery curriculum materials led to changes in the teaching of English?

Teachers from the intervention schools were asked to rate different units of the English Mastery curriculum materials, ranging from 1 ('terrible') to 5 ('excellent'). The first thing to note is that scores at the end of Year 7 for all six options were relatively high, particularly the curriculum map, which was the only option with a mean score above four. By the end of Year 8, the mean score for three of the six options increased slightly: units of work increased from 3.85 to 3.88, co-planning resources increased from 3.4 to 3.55 and knowledge definitions increased from 3.79 to 3.87. The remaining options decreased slightly, but their scores at the end of Year 8 were still relatively high. Overall, this suggests that teachers had a positive view towards the English Mastery programme materials throughout the trial.

Table 26: How would you rate the following English Mastery curriculum programme materials?

	End of Year 7			End of Year 8		
	N	Mean	SD	N	Mean	SD
Units of work	22	3.85	0.71	18	3.88	0.61
Lesson by lesson resources	22	3.93	0.70	19	3.86	0.72
Best practice videos	20	3.76	0.64	18	3.63	0.46
Co-planning resources	22	3.40	0.79	17	3.55	0.49
Knowledge definitions	22	3.79	0.75	17	3.87	0.60
Curriculum map	22	4.07	0.67	18	3.95	0.52

Source: Evaluation Teacher Survey. Responses: 1, 'terrible'; 2, 'poor'; 3, 'fair'; 4, 'good'; 5, 'excellent'.

Discrete grammar teaching, a pillar of Ark English Mastery Plus, was seen as an important element in improving teaching, particularly in terms of the emphasis on accuracy and grammatical terminology and how this could be embedded into the teaching of writing. English Mastery focused teachers on the explicit teaching of grammar and this was seen as an area that was previously lacking in the school English curriculum for some schools, and one that pupils entered secondary education with some confidence. In this way, leads often felt that the focus on grammar added continuity for pupils between KS2 and KS3.

'What I would say is that we have had to, well—how it's changed our existing practices I think is that there's a lot more emphasis on accuracy. The accuracy side of writing. That's how it's changed I would say' (Mastery Lead).

Furthermore, the focus on English Mastery was seen as building on what pupils had learned at KS2. This emphasis in turn influenced the focus and emphasis on explicit grammar teaching and its connection to the teaching of writing.

'I think the biggest impact for me ... has actually been on the staff. It's made staff really start thinking about how they are explicitly teaching grammar and how they are explicitly teaching writing and it's made us all more aware of maybe the gaps in other parts of our curriculum that we didn't even realise they were there' (Mastery Lead).

This was reflected in teacher comments about the technical aspects of writing and technical accuracy, such as punctuation and grammar, and now this, overall, gave them confidence to teach writing. When asked what had changed in their practice, teachers said:

'So it's made me more aware of the technical side of grammar and knowing how to answer questions' (teacher, Year 7).

'I think it's been very, very good at giving us a shared language, both in planning but also in delivery with the students in terms of grammatical things and drawing attention to SPaG and things like that' (teacher, Year 7).

'Oh definitely my focus on grammar, because on that lesson and other lessons there as well, a very specific focus on how the speech marks are used. A very particular focus on commas and different clauses, so almost that accuracy and grammar has changed' (teacher, Year 8).

The comments from teachers clearly indicate that their confidence to teach grammar and punctuation increased as a result of the programme. Given that improving subject knowledge with respect to grammar teaching was one of the most frequently cited reasons for joining the programme, and that grammar is integrated into the programme materials, it is not surprising that teachers focused on this in their responses when asked how their practice had changed.

While there is data suggesting that teacher confidence and subject knowledge for teaching English increased, two of the ten Mastery Leads interviewed perceived that the programme did not promote the enjoyment of English for all pupils. One teacher felt that their teaching was not 'as engaging as it could be' when following the programme and that their lack of enthusiasm about the programme meant that they were finding it difficult to be inspired when teaching it. A second teacher described how the programme felt 'repetitive' and that pupils found it 'monotonous'. That said, pupil focus-group data suggested that pupils were motivated by English Mastery and they enjoyed their Mastery lessons and had an increasing sense of self-efficacy (see section G).

e) Have English Mastery assessment protocols led to changes in teacher use of formative and summative assessment to inform planning and teaching?

Teachers from intervention schools were asked to indicate the extent to which English Mastery assessment protocols led to changes in the use of formative and summative assessment. A score of 1 represents 'no change at all' while a score of 5 represents 'complete change'. At the end of Year 7, the mean score was 3.26, however, this decreased to 2.81 by the end of Year 8 (see Table 27). A reason for this could be that by the beginning of Year 8 a number of schools had already changed their use of formative and summative assessments, resulting in less change by the end of Year 8.

Table 27: To what extent do you think that the English Mastery assessment protocols (for example, fortnightly quizzes and Assessing for Mastery standardisation) have led to changes in use of formative and summative assessment?

End of Year 7			End of Year 8		
N	Mean	SD	N	Mean	SD
21	3.26	0.78	19	2.81	0.87

Source: Evaluation Teacher Survey. Responses: 1 = 'not at all'; 5 = 'complete change'.

Teachers were also asked to rate different elements of the English Mastery curriculum integrated assessment model, ranging from 1 ('terrible') to 5 ('excellent'). At the end of Year 7 the scores for using formative and summative assessment to inform planning (3.63), marking English (3.62) and accurately assessing pupils' English (3.52) were relatively high, falling somewhere between the 'fair' and 'good' categories (see Table 28). These scores did reduce slightly by the end of Year 8 but they still remained within the same 'fair' and 'good' categories. It is difficult to explain the decline in scores here: the limited change and the small sample size mean that the results should be treated cautiously.

The embedded formative assessment opportunities were seen as beneficial by some teachers, particularly for building the confidence of pupils working at a lower level of ability:

'Lots of the English Mastery content has built-in opportunities like hinge questions and quizzes for formative assessment, which as a school we've been pushing anyway' (teacher, Year 7).

'So then the students are getting live feedback that they can respond to in the moment and they have a teacher there to help them understand the feedback' (teacher, Year 8).

With respect to termly assessment, while some teachers, after initially finding them challenging, grew in confidence in using them and felt that they were rigorous, others found them less useful for identifying the next steps in pupil learning as they lacked precision:

'I think that we feel confident in the sense that the system appears to be quite robust and thorough and it doesn't mean that it's not a challenge to engage with it, because it is quite challenging using standardisation materials and a very detailed mark scheme, but I would say that's a good challenge' (teacher, Year 7).

'There is a real rigor to that assessment now, which means that we can make more accurate judgements about where our students are. So I think that our confidence with that over the last year and a half has really, really grown' (teacher, Year 8).

'I really don't like them. They're too vague, they're not precise. They don't really have much weighting and I can't really gather from them what it is that the children are missing' (teacher Year 8).

Adherence to the assessment protocols of English Mastery was good, however a minority of schools did not follow this either due to time constraints or due to following a different system of assessment that was felt to be more beneficial.

Table 28: How would you rate the following elements of the English Mastery curriculum integrated assessment model?

	End of Year 7			End of Year 8		
	N	Mean	SD	N	Mean	SD
Accurately assessing pupils' English	22	3.52	0.85	20	3.38	0.79
Marking English	22	3.62	0.73	20	3.36	0.88
Using formative and summative assessment to inform planning	22	3.63	0.69	20	3.44	0.74

Source: Evaluation Teacher Survey. Responses: 1, 'terrible'; 2, 'poor'; 3, 'fair'; 4, 'good'; 5, 'excellent'.

f) What are teacher perceptions of the impact of English Mastery on the attainment and enjoyment of English for different groups of pupils?

Teachers from intervention schools were asked whether participation in English Mastery affected enjoyment in English for specific pupil groups, with responses on a scale from 1 ('slowed progress') to 5 ('accelerated progress'). Minor increases in enjoyment were observed for low attaining pupils, rising from 3.12 at the end of Year 7 to 3.32 by the end of Year 8, and SEN pupils, rising from 2.99 at the end of Year 7 to 3.2 by the end of Year 8. The only group with a drop in mean scores was FSM pupils (mean end Year 7: 3.34; end of Year 8: 3.18). However, it should be noted that perceived enjoyment at the end of Year 7 was relatively high compared to the other pupil groups and remained relatively high at the end of Year 8 (see Table 29).

There were slight perceived increases in progress in English across four of the five pupil groups. The largest increase in mean scores was observed for SEN pupils, rising from 3.03 at the end of Year 7 to 3.33 by the end of Year 8, followed by low attaining pupils, which rose from 3.11 to 3.36. The mean scores for FSM pupils were 3.29 across both years (see Table 30).

Table 29: How do you think that English Mastery has affected enjoyment of English for the following pupils?

	End of Year 7			End of Year 8		
	N	Mean	SD	N	Mean	SD
High attainment	22	3.06	1.21	20	3.03	1.27
Low attainment	22	3.12	1.14	20	3.32	1.00

SEN	22	2.99	1.18	20	3.20	0.99
FSM	20	3.34	0.89	20	3.18	1.03
EAL	21	3.05	0.86	20	3.08	1.02

Source: Evaluation Teacher Survey. Responses: 1, 'reduced enjoyment', to 5, 'improved enjoyment'.

Table 30: How do you think that English Mastery has affected progress in English for the following pupils?

	End of Year 7			End of Year 8		
	N	Mean	SD	N	Mean	SD
High attainment	22	3.19	1.11	20	3.26	1.17
Low attainment	22	3.11	1.23	20	3.36	0.93
SEN	22	3.03	1.07	20	3.33	0.89
FSM	20	3.29	0.99	20	3.29	0.90
EAL	21	3.16	1.02	20	3.22	0.89

Source: Evaluation Teacher Survey. Responses: 1, 'slowed progress', to 5, 'accelerated progress'.

Most teachers saw the overall programme as improving pupil confidence and progress in English. The quality of writing was also seen as improving. The content of the programme—such as all children accessing the same texts, Writing Mastery booklets, and the level of challenging but well-structured and scaffolded nature of the programme—was highly valued:

'I would say that particularly in terms of their confidence and their self-esteem, there seems to be a sense among the students that they feel more confident in what they're doing in English than I've seen historically' (teacher, Year 8).

'I think that the quality of their writing has improved through using the Writing Mastery booklets, definitely. I think that some pupils enjoy using the booklets and the scaffolding it gives for pupils to then get on to their writing and so I feel like pupils enjoy those parts' (teacher, Year 8).

'I think the ambition within the curriculum has been really important for pupils, that every single pupil—no matter what their ability is—has had the ability to access and to study these big canonical texts' (teacher, Year 8).

'I also think the kind of engagement with Shakespeare's language and not dumbing it down or using some kind of modern translation: I think that is really key, that the students are given that kind of academic rigor and challenge, which I can see how that has been built into the Ark curriculum' (teacher, Year 7).

Overall, these responses reflect aspects of the programme that teachers felt were impactful for pupil's writing and confidence.

A positive impact on reading comprehension related to the programme materials and level of challenges was further explained by one teacher:

'I think their comprehension has improved because the reading materials that we're using are... not only are they of a good quality and we really have the rigour needed, but there are glossaries through the reading section as well, so it helps to guide pupils' (teacher, Year 8).

g) What are pupil perceptions of the impact of English Mastery on their self-efficacy, confidence, and enjoyment in English?

Pupil focus-group data provided a combination of direct reports and analytic inferences about pupil perceptions of their English Mastery experience. Overall, this data suggested that English Mastery had some positive influence on pupil self-efficacy, confidence, and enjoyment of English. However, this impact varied across different elements of the programme. Key elements of the English Mastery approach that featured most prominently in pupil responses were:

- the study of texts;
- the teaching of vocabulary;
- the teaching of writing; and
- the role of the teacher—including their use of integrated assessment.

It is noteworthy that teachers were often present during the pupil focus groups, which may have encouraged pupils to respond positively. In addition, positive responses need to be viewed on the understanding that pupils are articulating their increased self-efficacy, confidence, and enjoyment within the parameters of the programme and its aims.

The study of texts

Teachers were mostly positive about the Literary Heritage strand and the embedded coherence of the programme. Likewise, most pupils reported enthusiastically on the texts they had studied as a class. They spoke up quickly and confidently in focus groups and their often-extended responses demonstrated an interest in, and understanding of, key features such as character relationships, analytical themes, and historical context. One pupil explained how reading *Animal Farm* had prompted him to start reading at home again:

'I mean, I don't read as much at home anymore, but I think I have, after we started reading Animal Farm I did get interested in some newer books that I have never read before' (pupil, Year 8).

Some pupils described these books as 'complex' but also as 'not childish', suggesting that studying Literary Heritage texts was confidence-building and competence-affirming. Other pupils spoke positively about how the organisation of the English Mastery programme enabled them to focus on one text at a time: 'Like last term we were doing *Oliver Twist* and now we're doing a play by Shakespeare.'

When asked, 'Are there any activities that you would like to do in your English lessons that you don't do at the moment?', some pupils commented that they would like greater opportunities to choose their own texts and have more time in the curriculum to read them. For example:

'Individual reading that you can go to the library and choose what book you want' (pupil, Year 8).

'Like more independent reading, instead of like whole-class reading, I think' (pupil, Year 7).

'Yeah, yeah, because we do library lessons sometimes. And we get to read a book of our choice. But I think it's important to be able to choose what we read as well sometimes during class' (pupil, Year 7).

Although these responses reflect varying perspectives and experiences, together they suggest a desire for greater autonomy and agency in aspects of text selection and reading engagement. It is also worth mentioning here that pupil perceptions of Reading for Pleasure were less well reported.

When pupils spoke about the teaching of texts in class, they often stated that they enjoyed tasks that involved active engagement and the application of existing knowledge. The study of plays proved particularly popular. When asked about activities that they would like to do in English lessons that they don't do at the moment, pupils wanted additional opportunities for practical engagement related to reading or acting out texts in class or watching a film adaptation or a stage production.

'I would like to do more, like, Shakespeare things as I think it would be fun—because last year I had a teacher that let us stand up and, like, read it out to each other and then we got to write about it and at the end of that type of topic we did like, we made our own version of it and wrote about it and described how the story went. It was really good' (pupil, Year 8).

Pupils' suggestions here also reflected a desire for activities that supported their comprehension of the text. For example:

'Yes, so when we do a Shakespeare play, we don't get to see the actual play and I feel it would be more helpful to see a performance of it done instead of just having the quotes and reading the script' (pupil, Year 8).

'I would probably watch, say if we're doing Shakespeare or something I was watching a play before I actually do the reading, so to help me understand' (pupil, Year 7).

The teaching of vocabulary

The explicit teaching of vocabulary also featured prominently in pupils' responses, reflecting this as a key element of the programme. Overall, the data suggested that pupils perceived the approach to teaching vocabulary positively.

'I don't know, but the hard words stand out. And they stand out more than words that we already understand, because if we don't understand a word then we're intrigued ... It's just interesting to learn what a word means' (pupil, Year 8).

Accumulating this new vocabulary was seen as an indicator of learning and a marker of their progress and achievement over time.

'My vocabulary has expanded more than it has before' (pupil, Year 8).

The pupils' frequent mention of 'vocabulary', particularly in connection with teacher assessment, conveyed a sense of confidence linked to achievement. In this way, the teaching of vocabulary appeared to have a similar positive impact on pupil self-efficacy to that of the literary heritage texts.

'It builds confidence, so any new words you learn. Because I didn't use as many words when I first came into the school. I've learnt loads more and how to use them' (pupil, Year 7).

The comment regarding confidence is an interesting one. It could suggest that this pupil understands that vocabulary is valued in school, or that they feel that they have greater tools for self-expression. Furthermore, some pupils talked enthusiastically about the word games they sometimes played in lessons and suggested they should play these more often.

'Like, usually people don't really like boring lessons, but if there is a game about it then they are more focused and engaged in it. Then they will probably learn more' (pupil, Year 7).

'I think, like, more interactive stuff... to allow everyone to join in' (pupil, Year 8).

These suggestions seemed to focus more on involving and engaging pupils practically in the lesson rather than on supporting their understanding. This may be an area for future consideration in the development of the programme.

The teaching of writing

In terms of Writing Mastery, pupil perceptions were mixed. For example, for some pupils the systematic teaching of grammatical concepts combined with deliberate practice was seen as beneficial and something they could apply in their own writing. This could reflect the ways in which pupils come to value what is presented as important to them in school. It could also reflect the teachers' growing confidence in teaching grammar to pupils effectively.

'I think the Mastery one is really good because you can understand more grammar and put it in your writing when you're writing essays and things' (pupil, Year 8).

'For me, in English I really didn't understand commas and how to use them, but finally after coming to secondary school and seeing that we have, like, whole lessons just about grammar and punctuation, it has

really helped me to kind of understand how to use commas in sentences and I can finally start understanding and putting them in my sentences' (pupil, Year 8).

However, pupil perceptions about Writing Mastery diverged sharply. Comments were often contradictory. For example, some pupils stated that they preferred Literary Heritage over Writing Mastery *because* they disliked the focus on grammar and punctuation in Writing Mastery—content which they felt they already understood and saw as repetitive.

'We take half the lesson to do a lot of questions but not a lot to write our opinions and what we want to write in Mastery lessons, because we do questions to understand the grammar and stuff but we sometimes go over things that we already know a lot of time, like commas, and using commas in the right places when most people do, and then not having enough time to write your ideas down' (pupil, Year 8).

In contrast, some responses suggested positive impact. This includes some Year 8 pupils reporting that Writing Mastery now made more sense to them than it did in Year 7.

'In Year 7 the layout of Writing Mastery is still the same, it's like the same books, but there are different things inside. Like in Year 7 I didn't really like Writing Mastery, but I've gotten to like it a bit more now and I can understand it now' (pupil, Year 8).

These examples are tentative indicators of a programme becoming more established in practice and of the teachers' familiarity and confidence with English Mastery increasing over time. However, a wide range of views were encountered in relation to Writing Mastery across Year 7 and Year 8 data with some pupils reporting engagement and others not.

Some pupils suggested that spending less time reviewing learning at the start of a lesson in Writing Mastery could free up more time for creative writing later on. Although some pupils talked about short story writing opportunities across genres (such as horror or comedy), others said that their lessons did not provide enough space for them to engage freely and creatively in writing. This suggested a desire for greater authorial agency. Some pupils wanted more opportunities to write freely in lessons in order to synthesise the different elements of the English Mastery programme they had been taught.

When asked, 'Are there any activities that you would like to do in your English lessons that you don't do at the moment?', the following responses were common:

'Probably more stories' (pupil, Year 7).

'Or maybe free-choice writing would be nice. Just every once in a while, do that' (pupil, Year 7).

'Yeah, I think that there should be more creative writing opportunities. I don't know how would fit that in, but I would appreciate more the ability to just show off your writing skills' (pupil, Year 8).

'I'd like more, so you can actually imagine things from your head and put it down on a piece of paper. Because just like [name] and [name] just said, we're usually just focusing on that one book. It's nice sometimes to do that, but sometimes you just want to just write down what you know and what you want to, a story or something like that' (pupil, Year 7).

The role of the teacher including their use of integrated assessment

Alongside these key elements, pupils frequently commented on the positive impact of the teaching by their English Mastery teachers. Pupils valued the specialist expertise and support of their teachers. They spoke about their teachers adapting lessons according to need, for example helping to make their lessons more interesting, more relatable, or more accessible. Above all, pupils identified the quality of their teachers' explanations as benefitting their understanding.

'I liked how she explained it. I think that she explained it in detail' (pupil, Year 8).

'Like, if we didn't get it, she would make it into something that teenagers might be able to get more' (pupil, Year 8).

'If the majority of people don't get the work, she pauses the class and then she brings us all together and the people that do get it explain, and then the people that don't get it understand and she explains further' (pupil, Year 7).

These quotes highlight the positive impact of English Mastery's integrated approach to assessment on pupil self-efficacy, confidence, and enjoyment. Further evidence was seen in relation to summative tests when, rather than feeling anxious about completing an assessment or being worried about being 'wrong', pupils often reported feeling confident and clear upon receiving regular positive reinforcement for their learning.

'I think it was very helpful. Like because there was an assessment coming up, it makes it easier for us to do well in an assessment' (pupil, Year 7).

'It makes me feel like I don't have to panic about it, and that I don't know what to do, but because we had that lesson I know what I should do now' (pupil, Year 7).

Overall, pupils talked positively about their perceived progress, next steps, and feedback. Pupils' examples tended to be writing-related. Writing conventions and technical accuracy (such as grammar, punctuation, paragraphs, spelling, and handwriting) were more evident than references to compositional choices or authorial voice. For example:

'I never knew, like, when we do paragraphs, I never knew—like it was ... I never knew about that. But then after Miss showed me it and like all that, every time we do quotes and things like that she always tells us 'Make sure to do that, makes sure to do that', to make the paragraphs better and my paragraphs have been getting better. And now she doesn't even need to tell me it anymore and I just know okay' (pupil, Year 8).

Summary of findings

Overall, these findings suggest that the English Mastery approach impacted positively on pupil self-efficacy, enjoyment, and confidence. This was most evident when the pupils were talking about the study of Literary Heritage texts, the teaching of vocabulary, tasks which prioritised active engagement and application, and the role of the teacher including teacher assessment. Pupils perceived their English lessons to be 'harder' at secondary school than they had been at primary school, but this challenge was seen as a positive one, with the pupil perception data conveying a sense of pride and achievement. Pupils cited the expertise and support of their English Mastery teachers as being significant in the context of the programme's teacher assessment strategies and key messaging, including around the importance of adapting teaching approaches and learning materials according to need.

The impact of Reading for Pleasure on pupil self-efficacy, enjoyment, and confidence was less well reported, which, alongside pupil perceptions on text choice and creative writing, suggested a perceived lack of pupil autonomy and agency. Pupil data revealed more about being a reader and less about being a writer. This suggests that authorship had less emphasis in the programme. The findings also suggested that an increase in opportunities for more personalised reading practices and more autonomous creative writing activities would address this, and lead to a more consistently positive impact on self-efficacy, enjoyment, and confidence overall.

In terms of explaining the discrepancy between the positive views of the intervention held by some participants and the absence of a positive effect in the impact evaluation, it is possible that school visits and survey responses included participants with more favourable experiences taking part. There is evidence from the compliance analysis that engagement was highly variable, which may have prevented any benefits from feeding through into pupil attainment within the study period.

h) What are school perceptions of the programme representing value for money?

The original intention was for the post-intervention teacher survey to include a question on school perceptions of programme value for money but in light of the response rates in the midpoint teacher survey, the evaluation team decided that collecting this data from a sample of schools would be likely to lead to findings biased toward schools that were more engaged with the programme and therefore more inclined toward favourable views. It is also worth noting that schools received the programme at a subsidised rate for this trial and the cost evaluation is based on figures that reflect what it

would cost a school ordinarily. The assessment of implementation costs presented below is therefore based solely upon figures provided by the delivery team at the end of the evaluation period.

Compliance

As discussed above, overall compliance was low, with only ten of the 30 intervention schools continuing in the trial until endline assessments being deemed compliant. Table 30 reports on monitoring data collected by the developer on whether intervention schools completed English Mastery activities sufficiently to be classed as compliant with the programme. This shows that most schools were able to meet the required threshold on four of the seven criteria. Of the 30 intervention schools, teachers from 27 attended induction training, 25 schools completed termly pupil assessments, 25 taught Literary Heritage and Mastery Writing for the recommended amount of time, and in 24 schools the English Mastery lead taught two lessons per week for at least two terms in both years of delivery.

However, for the other three indicators, fewer schools met the compliance threshold. Only 16 of the 30 intervention schools completed the recommended amount of co-planning. This is consistent with findings from the fieldwork, which showed that schools often struggled to find time for this even though it was seen as a valuable activity, as discussed in the IPE section. Furthermore, only half of intervention schools ($n = 15$) had their English Mastery lead attend a termly training session in at least two out of three terms in both years of the study, and the same number took up the support offered through four touchpoints each year. While the impact evaluation did not find any evidence of low compliance altering the overall trial results, the figures presented in Table 31 provide an insight into which elements of implementation proved challenging for schools and which might prove useful in considering how the programme and its expectations of participants might be refined. It is useful to note here that successful implementation of the English Mastery programme was contingent on structural support from senior leadership. In particular, this meant ensuring that there was sufficient timetabled lesson time to deliver English lessons and also sufficient time allocated to co planning and assessment.

Interviews with teachers gave some indication of difficulties they encountered in implementing the programme relating to coverage of the relevant units of work and finding the time to meet to co-plan and collaborate. Some expressed the view that there was too much content, with one specifically mentioning Literary Heritage and the difficulties of fitting in all curriculum content alongside the regular pupil assessments. Another suggested that their department may not have worked together enough to derive the maximum benefit from all elements of the programme. This may indicate that some schools found the full set of compliance measures as burdensome in terms of lesson and teacher time.

Table 31: Compliance with the intervention, for each indicator

Item	Compliance indicator	Overall threshold	N schools
1. Induction training attendance	KS3 teachers attend EM induction training.	80% of teachers delivering EM attend induction training.	27
2. Mastery Lead training attendance	EM Lead attends a termly training session (Mastery Lead workshop or Assessing for Mastery).	Attendance in at least 2/3 terms in Year 1 and in Year 2.	15
3. Co-planning time allocation	Co-planning allocated fortnightly at least ten weeks per term.	At least 2/3 terms in Year 1 and in Year 2.	16
4. Standardised conditions for assessment	Pupils sit the termly EM assessments in standardised conditions.	At least 2/3 terms in Year 1 and in Year 2.	25
5. English Mastery curriculum delivery	Schools teach Literary Heritage for 100 mins per week and Mastery Writing for 50 mins per week for at least ten weeks per term.	At least 2/3 terms in Year 1 and in Year 2.	25
6. English Mastery Lead involvement in teaching	EM Lead teaches at least two lessons per week for a minimum of ten weeks per term.	The English Mastery Lead teaches two lessons per week for 27 weeks of the year (70%).	24
7. Support uptake	EM Lead facilitates touchpoints with their school.	Four touchpoints each year: two school visits and two development sessions.	15

Fidelity

Interview data collected during school visits indicates that teachers and pupils clearly saw many benefits in English Mastery, however, implementation was time-intensive and often involved large changes to existing practice. Where this is the case, senior leadership and staff buy-in are critical to success of the programme. There were instances reported during field visits where staff turnover or changes to leadership meant that although the programme started well, adherence to the programme was difficult to manage over the two-year delivery period. This also included coverage of units of work. The amount of time scheduled to teach English often impacted negatively on coverage of the units (see Compliance section below). These factors explain why compliance to the intervention was low, as discussed in the Impact Evaluation Results section above. Lesson observations conducted by SHU English specialists during fieldwork visits to schools indicated that English Mastery lessons were taught using Ark materials. Levels of pupil engagement in the lessons observed was variable, and it is not possible to comment on whether this was a result of the programme or other factors such as the quality of teaching. However, variation in pupil engagement in lessons may suggest that the quality of implementation was not consistently high across all schools.

Usual practice

Interviews with Mastery Leads and teachers indicated that there was a range of practice taking place in English. Some intervention schools had previously followed an English Mastery programme and so adapted quite quickly to the Ark Plus model of English Mastery. Other schools had followed a much less structured programme for English where responsibility for planning was placed on the English lead and teachers. Some schools had previously worked from their own schemes developed by staff in the school. This has implications for how the programme was received by schools. For some it felt restrictive while to others the training, materials, and programme overall were appreciated as they added structure, coherence, and progression to the Year 7 and Year 8 English curriculum. While no data was collected that captured business as usual, schools were asked about their motivations for joining the programme. This revealed that concerns about KS3 coverage and progression to GCSE, and teacher subject knowledge, were key motivations for schools joining the programme. This suggests that schools perceived that usual practice was not addressing these factors.

Insights into usual practice in control schools are limited to the IPE survey questions on teacher workload presented in the first part of the IPE findings section. No fieldwork visits took place in control schools so there were no interviews with school staff or focus-groups with pupils through which to assess usual practice. Schools were not asked about their use of other relevant interventions during the recruitment process as the only eligibility criteria were that schools had pupils in the

relevant year groups and had not previously delivered English Mastery. Participation in other EEF trials during this evaluation was permitted although none of the schools visited mentioned this.

Cost

Schools pay an annual fee to deliver English Mastery. Different tiers of support are offered based on level of need. In this trial, the highest level of support, known as the 'tailored tier', was evaluated. A mid-tier option called 'leadership' is also available. Schools may access a further tier, with no school support, once they have completed three years of supported implementation. During the trial, schools subscribed to the tailored tier for two years at a subsidised cost of £900 per year. The most common model is for schools to implement the tailored tier for two years, then move to the leadership tier in the third year.

Programme pricing varies depending on school size, with slightly higher costs for larger schools. School size is calculated using pupil numbers from data on the Department for Education website. The costings below assume the school follows the most common journey: two years at the highest level of support followed by the mid-level of support in the third year. Prices shown are indicative for a standard-sized school. The programme also requires students to read set texts for the curriculum. The total cost assumes 120 students in a year group. Aside from the fees, schools must purchase the set texts taught as part of the programme. These costs are incurred annually and estimated at £16 per pupil per year as stated in Table 32. The developer does not require any additional expenditure on materials or equipment. Table 32 also displays the staff time required for English Mastery training and co-planning sessions. As the programme is delivered to pupils through classroom teaching, no further staff time is needed beyond that already allocated for English lessons.

Table 32: Cost of delivering English Mastery

Item	Type of cost	Cost	Total cost over 3 years	Total cost per pupil per year over 3 years
Annual licence fee, Tailored	Membership fee, year 1	£8,380	£8,380	
Annual licence fee, Tailored	Membership fee, year 2	£8,380	£8,380	
Annual licence fee, Leadership	Membership fee, year 3	£5,760	£5,760	
Set texts for Literary Heritage	Running cost per pupil	£16 per pupil per year	£5,760	
Induction training	Staff time	One day per teacher	-	
Mastery Lead training	Staff time	One day per term for Mastery Lead in each school	-	
Co-planning sessions	Staff time	One hour per week	-	
Total			£28,280	(£28,280/120/3) = £78.56

Table 33: Cumulative costs of English Mastery (assuming delivery over three years)

	Year 1	Year 2	Year 3
English Mastery	£10,300	£20,600	£28,280

Conclusion

Table 34: Key conclusions

Key conclusions	
1.	Pupils in English Mastery schools made no additional months' progress in English, on average, compared to pupils in other schools. This result has a low to moderate security rating.
2.	The trial aimed to evaluate the impact of English Mastery among children eligible for Free School Meals (FSM). There was a smaller number of pupils in this group and these results may have lower security than the overall findings. Among children eligible for FSM, those in English Mastery schools made two fewer months' progress in English, on average, compared to those in other schools.
3.	Compliance with the intervention was low, with only ten out of 30 intervention schools deemed fully compliant, meaning most pupils did not receive the programme as intended. There was a sense that the timetabling requirements for planning and training were demanding. Some schools reported that time constraints impeded successful implementation, including difficulties finding time to deliver full curriculum coverage and to schedule co-planning sessions, assessments, and training.
4.	Teachers suggested that English Mastery had a positive impact on their subject knowledge, particularly for teaching technical accuracy and grammar, and on workload.
5.	Teachers and pupils reported that English Mastery had a positive impact on pupil English skills and pupils reported increased self-confidence in this area. Teachers also reported that pupils benefited from the challenge of the Literary Heritage and Writing Mastery strands.

Impact evaluation and IPE integration

Evidence to support the logic model

The logic model for English Mastery includes outcomes for both teachers and pupils. With respect to teacher outcomes, it is clear from survey and interview data collected in the IPE that the training and materials improve teacher confidence in teaching canonical texts, grammatical concepts, and vocabulary. In addition, most teachers expressed more confidence in how to use assessment for planning, delivery, and effective feedback. They particularly appreciated the regular opportunities for formative feedback, finding that this motivated and supported pupils of lower ability. The logic model suggests that these elements will, in turn, improve pupil vocabulary and knowledge of grammatical concepts. Increased pupil self-efficacy, motivation, and engagement in English lessons were also intended outcomes of the programme. It is important to note here that the Ark Mastery programme was continually being developed and refined in response to teacher feedback. When comparing the results of Cohorts 1 and 2, this needs to be considered.

Some indicators of pupil impact such as progress to A-level were never going to be within scope for this trial due to the limitations of a two-year evaluation period. While there is no evidence of impact on pupil attainment in English as measured by the primary outcome, self-reported survey data from teachers suggests that there was some positive impact of the programme on both teachers and pupils. Survey and interview data suggest that the programme had some positive impact on teacher workload. In particular, teachers reported that they spent less time developing and planning resources. They valued the MyMastery platform resources, stating how these materials allowed them to spend more time adapting lessons to meet the needs of pupils, which impacted positively on the quality of teaching. This was mostly the case for teachers with less experience. The amount of experience held by participating teachers was highlighted as a contextual factor in the logic model, and findings from the fieldwork interviews showed variation in engagement and enjoyment according to how experienced teachers were.

Many teachers also told us that in terms of technical accuracy and grammar, their subject knowledge for teaching and understanding of how pupils progress in these areas improved throughout the programme, a finding consistent with teacher outcomes in the logic model. Most Mastery Leads and teachers found the use of standardised, norm-referenced assessments beneficial in identifying pupil strengths and areas for development. These assessments were integral to co-planning. However, as some schools found it difficult to schedule co-planning due to timetable constraints and competing priorities with their time, the impact of the programme's assessment, feedback, and planning cycle was limited.

While the Writing Mastery element of the programme was seen as conducive to progress in writing, some teachers and pupils raised concerns that pupils lacked the opportunity to write extended pieces of work or to write more ‘freely’ and have some ownership over the content of the writing tasks set. While there is clear evidence to suggest that pupils, particularly those who are less confident, require more concise tasks and frequent feedback to build their confidence, motivation, and self-efficacy, the role of extended writing and its value for all pupils is an area for future exploration. The relationship between pupil autonomy and writing and motivation and writing was not considered during this study but arose from the inductive coding of teacher interviews and pupil focus groups during field visits. Overall, both teachers and pupils agreed that the programme supported pupil-motivation through self-efficacy, and pupils enjoyed the challenge and depth of the Literary Heritage strand.

The findings suggest that embedding oracy into English lessons may be an important factor for both pupil motivation and engagement and also to enhance learning (see also EEF, 2021b). This arose when discussing EAL pupils, however, teachers said that most pupils benefit from an oracy focus in lessons or activities such as drama that make lessons more practical. Ark increased the focus on oracy during the programme as part of ongoing development of the materials and programme; this was very well received by teachers.

Interpretation

The headline finding from the impact evaluation is that the trial discovered no effect on the English attainment of participating pupils. While a degree of uncertainty exists around this headline result given that the trial was not powered to detect effects of such magnitude, there is no evidence that the intervention is associated with improved attainment as measured by the primary outcome. For one of the secondary outcomes, Reading, there is tentative evidence of one month less progress for pupils in intervention schools, although the range of possible results includes no effect; for the other secondary outcome, Grammar, Punctuation and Spelling, the point estimate indicates no effect. Two other impact evaluation findings are worth highlighting. First, a negative effect was reported for FSM pupils, who were estimated to make two months less progress than FSM pupils in control schools—although the range of possible effects includes no additional progress and the smaller sample size means this result is subject to considerable uncertainty. There is no evidence from the IPE that sheds any light on this finding and the results from the model including an FSM indicator in analysis of the whole sample were more similar to those obtained in the headline ITT model, again attesting to the uncertainty inherent in underpowered designs. Second, there was a pronounced difference in point estimates between the two delivery cohorts; this could be attributed to changes in the intervention for the second cohort, although the evaluation produced no evidence to confirm this.

Low compliance is one possible explanation for the lack of impact on pupil attainment. Two-thirds of intervention schools were not deemed compliant, including those that withdrew from the trial. This means that most pupils did not receive the programme as intended. While the analysis of compliance presented earlier in this report indicates that higher compliance may have resulted in a larger negative effect, using six compliance criteria demonstrates that effective implementation can be challenging and increasing this to seven criteria during the evaluation period could be seen as exacerbating this. The intervention essentially involves a radical overhaul of English teaching; this may be welcome in some schools but could conceivably deter others from taking part. When considered alongside the low compliance among intervention schools, this creates a picture of a programme that schools may find too demanding. This sentiment is likely reflected in the failure to reach the intended sample size, despite extending recruitment and delivery into a second year. The trial also suffered from school-level attrition with ten of the 72 schools withdrawing. Despite the positive feedback from many participants, this suggests that English Mastery can be difficult to deliver and that many schools may think that the comprehensive change required is not the right approach for them.

Limitations and lessons learned

There are several important limitations to the study design that should be acknowledged. Due to disappointing recruitment, the trial was split into two delivery cohorts, creating the possibility of differences between cohorts that could introduce confounding factors. Indeed, there were changes to the programme over the trial period and the impact evaluation results for the two cohorts were divergent. As the intended sample size was not achieved and attrition was substantial, statistical

sensitivity was weakened and the risk of bias increased. The lack of data on usual practice in control schools makes it difficult to know how different this was to intervention schools before or during the trial.

It is clear from this trial that the level and detail of specifications for programme fidelity are crucial, both in assuring adherence to criteria seen as central to the programme's efficacy and as providing a reasonable judgement of causality between the programme and its potential impacts. Where an intervention is both time intensive and involves substantial changes in school practice, it can be challenging to implement successfully. In this trial, we noted that for some schools the change to the English curriculum was substantial and involved the replacement of other practices in the teaching of English in both Year 7 and Year 8. For some schools this was welcome and met their improvement and development goals and motivations for joining the programme. For others, it replaced—or meant that there was insufficient lesson time to deliver—other aspects of English teaching. In future interventions, the scale of the change, and how far it impacts on teacher planning time and lesson delivery time, may need to be considered, along with embedding increased flexibility for schools.

The IPE findings point to key themes that are significant to teacher CPD. Most valued were the Mastery materials and school visits by delivery team members. The ongoing development of the programme in response to feedback from schools was another valuable element. Embedding support of this nature into all CPD programmes would be beneficial in implementing changing priorities of the curriculum. The individual school support, and suggestions of how the programme could be delivered successfully in unique contexts, was invaluable to teachers and leaders. Teacher workload is clearly a concern for schools and teachers. The English Mastery materials were welcomed as they freed teachers up to think more about how they would deliver the lessons, adapting them for their pupils according to their need. At different times in their career, teachers require a different type and level of support and professional development: CPD is most effective when tailored as such. Subject knowledge development seems a particular concern for less experienced teachers. Teachers commented positively on the challenge embedded in the Literary Heritage texts and the activities and lessons in the English Mastery planning resources. This was highly valued by both less experienced and more experienced teachers who were introduced to new ways of teaching complex and challenging texts. More experienced teachers tended to have an eye on pupil progression to GCSE English and reflected on how far the programme might support this. Gaining a picture of the overall progression towards GCSE English may be an essential consideration for staff development.

Future research and publications

According to teacher survey data, English Mastery strengthened teaching skills in writing, and also in reading comprehension, where immersion in a canonical text enabled pupils to develop their skills in reading and comprehending complex texts. Further exploration of the role of opportunities for extended writing and writing motivation would be a useful area for future exploration. The Mastery Writing focus on transcription skills meant that opportunities for extended writing and ownership over the writing process were often overlooked. This has implications for pupil authorship and autonomy over their writing.

Pupil focus groups are not commonly included in RCTs, however, in this trial, discussing with pupils provided insights that were useful in corroborating the perceptions of teachers and leaders. In addition to sharing their sense of emerging skills and confidence, it allowed pupils to suggest what else might be included in their English curriculum. Investigating the relationship between motivation and self-efficacy and opportunities for increased levels of writer autonomy would be a useful area for further investigation.

References

- Allen, R., Jerrim, J., Parameshwaran, M. and Thompson, D. (2018) 'Properties of Commercial Tests in the EEF database': https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/methodological-research-and-innovations/Research_Paper_1_-_Properties_of_commercial_tests.pdf
- Anders, J., Shure, N., Wyse, D., Barnard, M., Abdi, F. and Frerichs, J. (2021) 'Power of Pictures: Evaluation Report', London: Education Endowment Foundation.
- Bloom, H.S. (2006) 'The Core Analytics of Randomized Experiments of Social Research' (MDRC Working Papers on Research Methodology, August 2006): <https://www.mdrc.org/publication/core-analytics-randomized-experiments-social-research>
- Bloom, H. S., Richburg-Hayes, L. and Black, A. R. (2007) 'Using Covariates to Improve Precision for Studies That Randomize Schools to Evaluate Educational Interventions', *Educational Evaluation and Policy Analysis*, 29 (1), pp. 30–59.
- Boylan, M., Demack, S., Willis, B., Stevens, A., Adams, G. and Verrier, D. (2015) 'Multiplicative Reasoning Professional Development Programme: Evaluation': <https://www.gov.uk/government/publications/multiplicative-reasoning-professionaldevelopment-programme>
- Boylan, M., Wolstenholme, C., Demack, S., Maxwell, B., Jay, T., Adams, G. and Reaney, S. (2019) 'Longitudinal Evaluation of the Mathematics Teacher Exchange: China-England': <https://www.gov.uk/government/publications/evaluation-of-the-maths-teacher-exchange-china-and-england>
- Coldwell, M. and Maxwell, B. (2018) 'Using Evidence-Informed Logic Models to Bridge Methods in Educational Evaluation', *Review of Education*, 6 (3), pp. 267–300.
- Cremin T., Mottram, M., Collins, F., Powell, S., and Safford, K. (2014) *Building Communities of Engaged Readers: Reading for Pleasure*, London: Routledge.
- Culliney, M., Daniels, K., Booth, J., Coldwell, M., and Demack, S. (2021) 'REACH Primary Evaluation Report', London: Education Endowment Foundation: <https://d2tic4wvo1iusb.cloudfront.net/>
- Davies, M., Read, H. and Guscott, E. (2022) 'English Mastery Evaluation Report', London: Education Endowment Foundation: <https://d2tic4wvo1iusb.cloudfront.net/documents/projects/English-Mastery-Final.pdf?v=1651210865>
- Demack, S. (2019) 'Does the Classroom Level Matter in the Design of Educational Trials? A Theoretical and Empirical Review', London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/public/files/Publications/Does_the_classroom_level_matter.pdf
- Demack, S., Boylan, M., Culliney, M. and Wolstenholme, C. (2022) 'Evaluation of Realistic Maths', London: Education Endowment Foundation: <https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/realistic-maths-education>
- DfE (2012) *Research Evidence on Reading for Pleasure*, Crown Copyright.
- Dracup, T. (2014) 'The Politics of Setting': <https://giftedphoenix.wordpress.com/2014/11/12/the-politics-of-setting/>
- Dong, N., Kelcey, B., Maynard, R. and Spybrook, J. (2015) *PowerUp! Tool for Power Analysis*.
- EEF (2021a) 'Metacognition and Self-Regulated Learning: Guidance report', London: Education Endowment Foundation: <https://educationendowmentfoundation.org.uk/education-evidence/guidance-reports/metacognition>
- EEF (2021b) 'Improving Literacy in Secondary Schools', London: Education Endowment Foundation: <https://educationendowmentfoundation.org.uk/education-evidence/guidance-reports/literacy-ks3-ks4>
- Guthrie, J. T. and Wigfield, A. (2000) 'Engagement and Motivation in Reading', in M. L. Kamil, P. B. Mosenthal, P. D. Pearson, and R. Barr (eds.), *Handbook of Reading Research (3rd edn)*, New York: Longman.
- Guthrie, J. T., Hoa, L., Wigfield, A. and Tonks, S. (2007) 'Reading Motivation and Reading Comprehension Growth in the Later Elementary Years', *Contemporary Education Psychology*, 32 (3), pp. 282–313.

- Hempel-Jorgensen A., Cremin, A. T., Harris, D. and Chamberlain, L. (2018) 'Pedagogy for Reading for Pleasure in Low Socio-Economic Primary Schools: Beyond "Pedagogy of Poverty"?' *Literacy*, 52 (2), pp. 86–94.
- Kelcey, B., Spybrook, J., Phelps, G., Jones, N. and Zhang, J. (2017) 'Designing Large Scale Multisite and Cluster Randomized Studies of Professional Development', *Experimental Education*, 85 (3) pp. 389–410.
- Maxwell, B., Daniels, K., Demack, S. and Stevens, A (2018) 'Mastery English and Maths', Learning Unlimited Teaching School Alliance (LUTSA): Strategic School Investment Fund (SSIF) Round 1 Project Evaluation.
- Ng, C. and Graham, S. (2018) 'Improving Literacy Engagement: Enablers, Challenges and Catering for Students from Disadvantaged Backgrounds', *Research in Reading*, 41 (4), pp. 615–624.
- Sedgewick, E. (2001) *Teaching Literacy: A Creative Approach*, London: Continuum.
- Wentzel, K. and Wigfield, A. (1998) 'Academic and Social Motivational Influences on Student's Academic Performance', *Education Psychology Review*, 10 (2), pp. 155–175.

Appendix A: EEF cost rating

Figure 2: Cost Rating

Cost rating	Description
£££££	Very low: less than £80 per pupil per year.
£££££	Low: up to about £200 per pupil per year.
£££££	Moderate: up to about £700 per pupil per year.
£££££	High: up to £1,200 per pupil per year.
£££££	Very high: over £1,200 per pupil per year.

Appendix B: Security classification of trial findings

Appendix: Padlock rating

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	MDES	Attrition					
5	Randomised design	≤ 0.2 (=0.19)	0-10% (=9%)					
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%					
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%	3		 Adjustment for threats to internal validity [-1]		
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					2
0	No comparator	≥ 0.6	$>50\%$					

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Low	RCT design used, controlling for observable and unobservable differences, and allocation to trial arms conducted independently. School characteristics broadly balanced across arms. Slight imbalance in baseline measures used for primary analysis, but less than 0.05 and controlled for in the regression model.
Threat 2: Concurrent Interventions	No information available/ Potentially moderate	No data were collected to assess the control schools' offer, an important limitation of the evaluation. Motivation to join the trial, collected from treatment schools, suggests that schools sought support for English teaching; therefore, there might have been overlapping, unknown English-teaching practices across schools, which could have influenced the outcome.
Threat 3: Experimental effects	No information available/ Potentially Moderate	No evidence to assess this. As per the rationale set out in the threat above, it could have been the case that schools allocated to control sought similar programmes and decided to implement them (compensatory rivalry).
Threat 4: Implementation fidelity	Moderate/High	Compliance was well-defined and aligned with the logic model. However, compliance was low (only 10 of 30 schools were deemed compliant), which may dilute the effect.
Threat 5: Missing Data	Moderate	Missing data is moderate and quite balanced across arms: 27.6% attrition in intervention group and 26.6% attrition in control group. Multiple imputation was conducted, and did not substantially alter the results of the complete case analysis.
Threat 6: Measurement of Outcomes	Low	Externally, reliable and validated commercial measure used for post-testing, which were externally marked (blinded to allocation). Reliable, valid tests were used at baseline (NPD KS2 GPS and KS2 reading) as well.
Threat 7: Selective reporting	Low	No evidence of selective reporting. The study is registered and the protocol is published. Deviations from the original protocol are reported in an updated version.

- **Initial padlock score:** [3] Padlocks – MDES at randomisation was 0.21 (-1 padlock) and attrition was 27.6% (intervention group) and 26.6% (control group) between randomisation and analysis (-1 padlock). Thus, two padlocks dropped as per guidance
- **Reason for adjustment for threats to validity:** [1] Padlock – two moderate threats were identified. Implementation fidelity is likely to underestimate the intervention's observed impact. The missing data has

an unknown direction of effect. It was also not possible to accurately assess two of the potential threats to validity (concurrent interventions and experimental effects) due to insufficient information. Theoretically, they could have affected the intervention's impact – hence dropping one instead of zero padlocks.

- **Final padlock score:** initial score adjusted for threats to validity = [2] Padlocks
- **Reason(s) for final padlock rating:** Initial score of three padlocks, minus one padlock due to identified threats to validity leaves a final padlock rating of two padlocks.

Appendix C: Effect size estimation

Appendix Table 2: Effect size estimation

			Intervention group		Control group			
Outcome	Unadjusted differences in means	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance	Population variance (if applicable)
GL PTE13	-0.86	-0.04	1824 (501)	194.64	1956 (651)	195.64	195.41	

Further Appendices

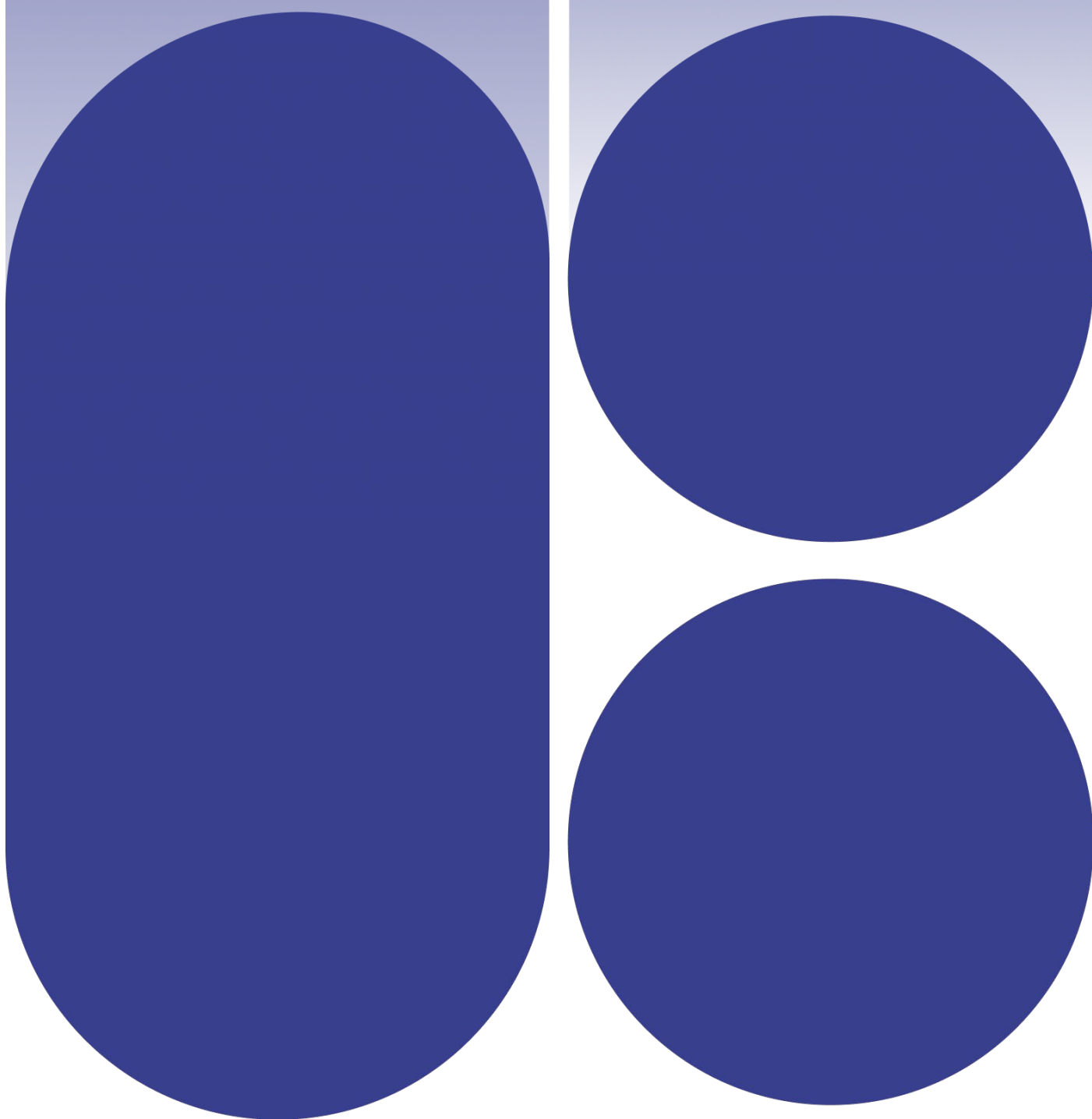
Appendix D – H can be found published on the project page labelled as '*Further Appendices*'.

You may re-use this document/publication (not including logos) free of charge in any format or medium under the terms of the Open Government Licence v3.0.

To view this licence, visit <https://nationalarchives.gov.uk/doc/open-government-licence/version/3> or email: psi@nationalarchives.gsi.gov.uk

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

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



**Education
Endowment
Foundation**

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

<https://educationendowmentfoundation.org.uk>

 @EducEndowFoundn

 [Facebook.com/EducEndowFoundn](https://www.facebook.com/EducEndowFoundn)