Independent evaluation of the Mastering Maths programme: a two-arm cluster randomised trial
Evaluation Protocol

Evaluator (institution): NatCen
Principal investigator: Helena Takala

Evaluation summary

Project title	**Independent evaluation of the Mastering Maths programme: a two-arm cluster randomised trial**
Developer *(Institution)*	University of Nottingham (UoN)
Evaluator *(Institution)*	National Centre for Social Research (NatCen)
Principal investigator(s)	Helena Takala
Protocol author(s)	Helena Takala, Enes Duysak, Abigail Rennick, Anjhana Damodaran, Tien-Li Kuo, Charlotte Tomlinson
Trial design	Two-arm cluster randomised trial
Trial type	Effectiveness trial
Pupil age range and Key stage	Students aged 16-19
Number of teachers *(at design stage)*	140 teachers in Further Education (FE) settings, with a maximum of two teachers eligible from a single college setting
Number of students *(at design stage)*	7,000 students
Primary outcome measure and source	GCSE Maths score, collected directly from FE settings
Secondary outcome measure and source	GCSE Maths score and grade, collected directly from FE colleges Self-confidence and self-efficacy in maths, measured with University of Manchester's self-efficacy questionnaire and the Attitude towards Maths Inventory (ATMI)

Protocol version history

Version	Date	Reason for revision
1.0 **[*original*]**	04/09/2024	N/A
2.0 **[first revision]**	02/02/2026	This version of the protocol includes updates to the impact evaluation design agreed when drafting the statistical analysis plan. These include: • Updating the stratification variables used for randomisation, removing group allocation in the efficacy trial to leave only FE setting • Removing a covariate from the primary analysis model – binary indicator of whether a student took the November resit • Adding a potential alternative source for endline data collection – exam boards • Changing the National Pupil Database variable used for the baseline measure of KS2 maths attainment, from 'KS2_MATHSCORE_noSpeccon' to 'KS2_MATMRK' • Adding two subgroup analyses, one for an alternate measure of disadvantage (residing in one of the 27% most deprived areas, according to the Index of Multiple Deprivation 2019) and another for low prior attainment (defined as having previously received a grade 2 or below in GCSE Maths) • Updating the framing of the research questions, including two new questions for the added subgroup analyses and splitting the question on attitudes towards maths in two: one question on self-confidence in maths and another on self-efficacy in maths • Including the details of cognitive testing undertaken to pilot the secondary outcome measures for self-confidence and self-efficacy in maths

		• Adding the 'without attrition' power calculation scenario for FSM eligible students to Table 2 • Adding two sensitivity analyses, one to assess whether November resit exams influence the results and another to support the EEF's synthesis work by replicating the primary analysis model including only prior attainment and the stratification variable • Updating the approach to mediation analysis, to specify that self-confidence and self-efficacy will be combined into a single mediator representing attitudes towards maths This version of the protocol also includes agreed updates to the IPE and cost evaluation design. These include: • Adjusting which IPE dimensions were captured by each IPE activity • Updating the approach for case studies to be able to explore contamination between intervention and control teachers • Updating the approach for student interviews to support recruitment and data quality • Updating the approach for survey administration due to changes in timeline and better-then-expected response rates from teachers • Changing the data sources for the cost evaluation, removing lead teacher interviews in favour of a cost proforma completed by the delivery team at endline Finally, we have corrected typos and clarified meaning in places throughout this version of the protocol, included the trial registration link on page 43, updated the details of the evaluation team and updated the timeline to reflect changes to data collection windows.

Contents

Tables

Figures

Study rationale and background

Policy context

Numeracy and mathematical skills are key to success in entering and progressing within the UK labour market. The Department for Education's (DfE's) Review of post-16 mathematics in 2017 highlighted that adults with basic numeracy skills go on to earn higher wages and are more likely to be employed than those without. Higher levels of mathematical skills are also associated with increased productivity and therefore economic growth. Despite increasing demand in the labour market for mathematical skills, particularly due to advancements in technology which are changing the nature of work, there is a persistent shortfall in supply. England has lower maths achievement levels for post-16 students compared to other developed countries. It was also the only country in a 2013 sample where young adults did not perform better than older adults in numerical proficiency. The review concluded that there is a need for further ongoing investment to improve mathematics teaching in Further Education (FE) colleges[1], which is where over a third of post-16 students study.[2]

Following the 2011 Wolf review[3], which outlined recommendations for how to improve vocational education, the government introduced a condition of funding.[4] Since 2015, all students aged 16-19 who have not achieved a grade 4/C in maths and English have to continue studying the subjects until they meet the required grade. The aim of the policy is to improve maths and English skills of students, to better equip them for the labour market and for life. The results for maths have been mixed; while the number of students resitting their maths exam has increased since the introduction of the policy, from approximately 100,000 in 2014 to 154,023 in 2023, the number of those achieving a grade 4/C during resits has declined, from approximately 40,000 in 2015 to 25,000 in 2023. [5,6] It has also been documented that each time a student resits, they have a lower chance of achieving a grade 4/C.[7]

Since the introduction of the policy, there have been a number of studies published on maths in the FE context. They suggest that the learning experience for post-16 students in FE needs to be different in comparison to younger students at school, especially by being tailored and appropriate to their needs. In particular, these students may have an "insecure understanding of basic [mathematical] concepts"[8] and may therefore benefit from alternative teaching strategies. The same students can also feel a lack of motivation towards maths, often linked to negative previous experiences of the subject.[9]

---

[1] Smith. (2017). Report of Professor Sir Adrian Smith's review of post-16 mathematics.
[2] Association of Colleges. 2020. College Key Facts 2019/20.
[3] GOV.UK Independent report. 2011 Review of vocational education: the Wolf report.
[4] GOV.UK Guidance. 2024. 16 to 19 funding: maths and English condition of funding.
[5] The Royal Society. 2022. GCSE Resits Project: Background.
[6] Camden, B. 2023. GCSE resits 2023: Maths and English pass rates down again.
[7] House of Commons Education Committee. 2023. Oral evidence: Government Proposals for Compulsory Maths to Age 18.
[8] Wake, G., et al. 2023. Centres for Excellence in Maths Teaching for Mastery Randomised Controlled Trial, Evaluation Report.
[9] Noyes & Dalby. 2020. Mathematics in England's Further Education Colleges: an analysis of policy enactment and practice

Notably, 43% of those teaching GCSE Maths in these contexts do not have a maths degree or an A-Level (or equivalent) in maths.[10] There is also limited support and few professional development opportunities for those teaching post-16 maths resit students.[11] To address this, academics at the University of Nottingham (UoN) developed a continuing professional development (CPD) programme for teachers, Mastering Maths, to support the need for upskilling teachers in the FE sector and in turn improve the number of students achieving a grade 4/C or above. The approach aims to address the possibility that students resitting their GCSE Maths may not have a secure understanding of basic mathematical concepts and therefore works to develop this, in addition to their fluency, while ensuring to value and build upon their prior learning.[12]

Evaluation evidence

Mastering Maths has previously been evaluated for efficacy. The previous evaluation was funded by the DfE as part of the Centres for Excellence in Maths (CfEM) programme, managed and led by the Education and Training Foundation (ETF), and implemented by UoN in the 2021-2022 academic year.[13] The three-armed cluster-randomised controlled efficacy trial involved 147 FE settings and 7,453 students. Key findings suggested that GCSE resit students taught by teachers in the full teaching intervention made one month's additional progress in maths learning compared to similar students in other teaching-as-usual settings. Students who had been eligible for free school meals (FSM) prior to joining FE settings and who were taught by teachers in the full intervention, made two months' additional progress in maths learning compared to students in teaching-as-usual settings. UoN hypothesised this was due to these students benefitting the most from the classroom being a safe and social space, and attention being paid to their prior learning enabling this to be built upon.

The efficacy trial was hindered by the COVID-19 pandemic, with high attrition rates among teachers and disruption to teaching. As a result, the current study aims to build on the evidence found during the smaller scale efficacy trial by conducting an effectiveness trial, which will aim to establish the effect of the intervention in more typical conditions. This will be achieved by randomly assigning teachers to either receive the Mastering Maths intervention or teaching-as-usual (control) and comparing the difference in the attainment of students of teachers in the two groups. In addition, it will examine the effect of the intervention when delivered using a more scalable model, whereby lead teachers provide professional development to GCSE teachers, rather than this being delivered by the developers. This is referred to as 'Train the Trainer'. For an overview of differences between the two trials, please see Appendix A.

Integrated evaluation design

This evaluation of Mastering Maths will consist of an impact evaluation (IE) and an implementation and process evaluation (IPE) which will run over two years, from 2024 to 2026. The aim of the IE will be to ascertain whether Mastering Maths improves students' GCSE Maths performance, in addition to exploring whether their attitudes towards maths improve over time,

[10] Maughan, S., et al. 2016. Improving Level 2 English and maths outcomes for 16 to 18 year olds – Literature review.
[11] Crips, B., et al. 2023. Post-16 GCSE Resit Practice Review. Centre for Education & Youth, and University of Warwick.
[12] Education & Training Foundation. Centres for Excellence in Maths Mastery Principles.
[13] Wake, G., et al. 2023. Centres for Excellence in Maths Teaching for Mastery Randomised Controlled Trial, Evaluation Report.

and whether the impact differs for students from disadvantaged backgrounds, as measured by prior FSM status up to age 16. The IPE will aim to understand how the intervention was delivered in practice, what the perceived impacts were on participants, including for disadvantaged students, and the relationships which exist between outcomes and moderators/contextual factors.

Intervention

Name

Mastering Maths

Who (recipients and provider)

The developers of Mastering Maths based at the UoN, herein referred to as the delivery team, will recruit and provide training to a group of lead teachers. Lead teachers will then proceed to implement a professional development programme to intervention teachers at FE colleges across England. Following this, these teachers participate in lesson study groups hosted by their lead teacher, with a small group of other participating teachers. All participating teachers teach five Mastering Maths lessons to 16-19-year-old GCSE Maths resit students spaced over the period November to March 2024-5.

What (materials and procedures)

Mastering Maths is underpinned by five key principles.[14] These principles are:

- **Developing an understanding of mathematical structure**. Representations can be used to unlock understanding so that students know the 'why' and not just the 'how'. They can both clarify the meaning of a concept and provide access to the structure of mathematical problems.
- **Value and building on students' prior learning**. Celebrating what students already know and making maximum use of the teaching time available to fill in key gaps in knowledge and understanding. There may be a number of misconceptions that need to be unpicked.
- **Prioritising curriculum coherence and connections**. Encouraging students to see the links between mathematical concepts (for example, similarity, ratio and trigonometry) rather than seeing them as separate content that need to be individually learned.
- **Developing both understanding and fluency in mathematics**. Covering key content in depth to attain fluency and understanding that can be applied in different contexts is preferable to superficial coverage of a large amount of material. Fluency is not just about knowing facts and procedures, but also how and when to use them.

[14] Education & Training Foundation. Centres for Excellence in Maths Mastery Principles.

- **Developing a collaborative culture in which everyone believes everyone can succeed**. The belief that effort leads to improvement. Low-threat/high-challenge tasks and activities.

The delivery team at UoN will train the lead teachers, to enhance their understanding of the Mastering Maths approach and the process of lesson study. Lead teachers will take part in five days of professional development during which they will develop a detailed understanding of the design of the twelve research lessons, the process of lesson study and the design and intentions of the professional development programme for the intervention teachers. Lead teachers will receive materials to support them in running the intervention teacher professional development and the lesson study cluster meetings.

Following this, lead teachers will implement two days of professional development for the intervention teachers, to enhance their understanding of Mastering Maths teaching approaches. The professional development will be run in regional clusters in London, Sheffield, Manchester and the South West. The intervention teachers will be given a handbook titled 'Mastery Teaching in Further Education' and signposted to the twelve exemplar Mastering Maths lessons with research questions and student tasks, available on the ETF website.

Lead teachers will also run five lesson study group sessions to a small group (approximately five) of the intervention teachers. One teacher at a time will volunteer to host a lesson study group at their setting, during which they will teach a Mastering Maths lesson to students while the other teachers and lead teacher observe. Afterwards, the lesson study group will meet to discuss the Mastering Maths teaching approaches in the context of the particular lesson, using a set of research questions to focus their conversation.

Each group of teachers will teach and use the same lesson (Lesson 2) for their first lesson study. As a group, they will choose which four of the remaining eleven lessons they will use to guide their work for the other sessions. The groups will self-organise who will host and teach a particular lesson. It is intended that prior to the lesson study, all teachers will already have taught the lesson to their GCSE classes, thus increasing familiarity.

How

All Mastering Maths activities will take place in-person. This includes the training to lead teachers, the professional development for the intervention teachers, the lesson study groups, and the Mastering Maths lessons taught by intervention teachers.

Where

The intervention will be run in FE colleges across England.

When and how much (dosage)

UoN will provide five days of training to the lead teachers in June and September 2024. Subsequently, the lead teachers will provide two days of personal development to the intervention teachers in October 2024.

From November 2024 to March 2025, teachers in the intervention group will teach five Mastering Maths lessons to all their GCSE resit class. This will be approximately one lesson per month. They will also take part in five lesson study groups with the same frequency.

Tailoring (adaptation)

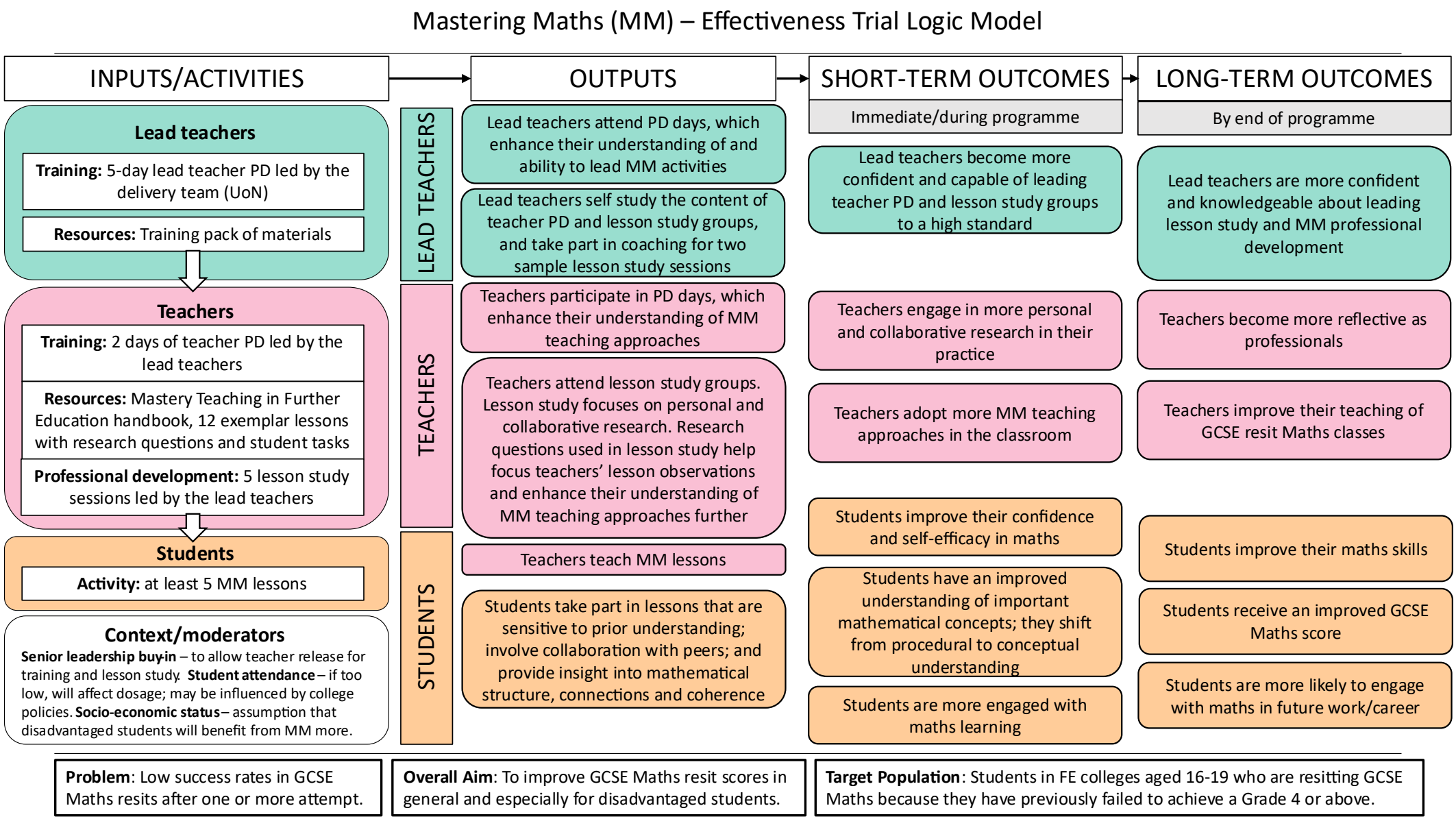
Teachers have some flexibility over how, as a small team working with a lead teacher, they will engage with the Mastering Maths lessons. Each team will teach and use Lesson 2 for their first lesson study. As a group, they will then choose which four of the remaining eleven lessons they will collaboratively work with over the course of the programme. They will also be able to choose who will host and teach the chosen lesson, in the five lesson study group sessions. The choice and flexibility are designed to ensure that the lessons best fit the needs of teachers in each group and to encourage teachers to engage in personal and collective research of their practice.

Lead teachers and intervention teachers are given a script to follow for the lesson study group meetings and Mastering Maths lessons, respectively. Lead teachers and intervention teachers will both be able to adapt the materials, including the script, to ensure that implementation of training and teaching best responds to teachers and their students as learners. Lead teachers will support their teachers to ensure good levels of compliance to the programme.

Logic model

Building on the logic model developed for the efficacy trial, we developed an updated logic model for the Mastering Maths effectiveness trial in collaboration with UoN, in Figure 1.

Figure 1: Logic model



Impact evaluation design

Research questions

The research questions which the impact evaluation of Mastering Maths aims to answer are the following: [15]

RQ1: What is the impact of Mastering Maths on 16- to 19-year-old students' GCSE Maths scores relative to those students receiving teaching-as-usual? (primary outcome)

RQ2: What is the impact of Mastering Maths on the probability of students achieving a grade 4 or higher on their GCSE Maths resit relative to those students receiving teaching-as-usual? (secondary outcome)

RQ3: What is the impact of Mastering Maths on the probability of students moving up at least one grade on their GCSE Maths resit relative to those students receiving teaching-as-usual? (secondary outcome)

RQ4: What is the impact of Mastering Maths on students from disadvantaged backgrounds as measured by prior FSM status relative to those students receiving teaching-as-usual? (secondary outcome)

RQ5: What is the impact of Mastering Maths on students from disadvantaged backgrounds as measured by the Index of Multiple Deprivation relative to those students receiving teaching-as-usual? (secondary outcome)[16]

RQ6. What is the impact of Mastering Maths on students with lower prior attainment relative to those students receiving teaching-as-usual? (secondary outcome)[17]

RQ7: What is the impact of Mastering Maths on students' self-confidence in maths relative to those students receiving teaching-as-usual? (secondary outcome)

RQ8: What is the impact of Mastering Maths on students' self-efficacy in maths relative to those students receiving teaching-as-usual? (secondary outcome)

Design

Table 1: Trial design

Trial design, including number of arms	Two-arm cluster randomised controlled effectiveness trial
Unit of randomisation	Teacher level

[15] The research questions were reformulated from their original phrasing ("Does Mastering Maths improve/increase …") to adopt an impact-oriented framing ("What is the impact …").

[16] The funding formula for all providers delivering 16-19 education includes a disadvantage funding element based on students' economic deprivation (via Index of Multiple Deprivation) as well as low prior attainment. Details are covered in the Secondary outcome section.

[17] During the peer review of the Statistical Analysis Plan (SAP), a subgroup analysis based on prior attainment was introduced and subsequently agreed to be articulated as an independent research question.

Stratification variables		Further education (FE) setting[18]
Primary outcome	**Variable**	RQ1 – GCSE Maths score
	Measure *(instrument, scale, source)*	RQ1 – GCSE Maths standardised raw score (z-score by exam board), collected directly from FE colleges or exam boards
Secondary outcome(s)	**Variable(s)**	RQ2 – Probability of achieving a grade 4 or higher in GCSE maths RQ3 – Probability of moving up at least one grade in GCSE Maths RQ4-6 – GCSE Maths score RQ7 – Self-confidence in maths RQ8 – Self-efficacy in maths
	Measure(s) *(instrument, scale, source)*	RQ2 – GCSE Maths grade, collected directly from FE colleges or exam boards, a binary variable equal to 1 if GCSE Maths grade is 4 or higher and 0 if it is 3 or below RQ3 – GCSE Maths grade, collected directly from FE colleges or exam boards, a binary variable equal to 1 if student moves up a GCSE Maths grade and 0 if not RQ4-6 – GCSE Maths standardised raw score (z-score by exam board), collected directly from FE colleges or exam boards RQ7 – Attitude towards Maths Inventory (ATMI) anglicised version, Likert scale (1-5) RQ8 – University of Manchester's self-efficacy questionnaire, confidence segment, Likert scale (1-4)
Baseline for primary outcome	**Variable**	Key Stage 2 (KS2) maths attainment
	Measure (instrument, scale, source)	KS2 National Curriculum Test, Total marks achieved in maths test ('KS2_MATMRK' variable, 0-110), National Pupil Database (NPD)[19]
	Variable	KS2 maths attainment

[18] Initially, we planned to stratify by both setting and group allocation in the efficacy trial (i.e., whether participants were in the full intervention or not). However, since none of the teachers participated in the efficacy trial, stratification was limited to setting only. The evaluation protocol has been updated to reflect this change.

[19] We initially planned to use the Maths scaled score (KS2_MATHSCORE_noSpeccon variable, 0-120) due to uncertainties surrounding the availability of raw maths attainment scores in the NPD data. However, after communicating with the Department for Education, we confirmed that the raw scores, labelled as 'marks', are available in the NPD. We then choose using the raw score ('KS2_MATMRK' variable) as a baseline measure, in line with the EEF's statistical guidance (2022).

Baseline for secondary outcomes	**Measure** (instrument, scale, source)	KS2 National Curriculum Test, Total marks achieved in maths test ('KS2_MATMRK' variable, 0-110), National Pupil Database (NPD)

The evaluation will be conducted as a two-arm cluster randomised controlled effectiveness trial of the Mastering Maths programme. We will build this evaluation based on findings from the efficacy trial and assess the real-world impact of the Mastering Maths programme beyond controlled and ideal conditions. The efficacy trial tested two versions of the Mastering Maths programme and showed that the full-intervention (with three-days professional development, lesson study groups and teaching materials) was the most effective version.[20] This finding informed which version of the programme would be tested for the effectiveness trial.

The randomisation in this trial is at the teacher-level, with stratification by FE setting.[21] A setting is defined as a group of GCSE Maths resit teachers that are connected due to the organisational structure and/or geographic location of their FE college. The teachers who agree to participate in this trial will be randomised to either the Mastering Maths group or the teaching-as-usual group, with a 50:50 ratio of teachers between the two groups.

The primary outcome of interest is the GCSE Maths scores of students resitting the exam (RQ1). The GCSE Maths raw score will be standardised by exam board and will be directly obtained from FE colleges. The secondary outcome will use students' GCSE Maths grade to estimate the probability of students achieving a grade 4 or higher on their GCSE Maths resit (RQ2), and the probability of students moving up a grade on their GCSE Maths resit (RQ3). We will also use GCSE Maths scores to assess whether the impact of Mastering Maths differs for students from disadvantaged backgrounds, as measured by prior FSM status (RQ4) and the Index of Multiple Deprivation (RQ5). In addition, we will assess whether the impact differs for students with low prior attainment (RQ6).[22] To estimate whether Mastering Maths improves students' attitudes towards maths, we will measure students' self-confidence (RQ7) and self-efficacy (RQ8) using the anglicised version of the self-confidence subscale of the Attitude towards Maths Inventory (ATMI) and University of Manchester self-efficacy questionnaire. For all outcome measures, KS2 Maths attainment score from NPD will be used as the baseline measure.

Cognitive testing

As the self-confidence and self-efficacy scales have not previously been used with this specific student population, we will conduct ten cognitive interviews with students currently retaking their GCSE maths, aged 16-19 and based in FE colleges and sixth-form colleges in England. They will be screened to ensure they are not based at a college taking part in the main Mastering Maths trial. The main aim we will seek to answer through the cognitive fieldwork is whether the questions

[20] Wake, G., et al. 2023. Centres for Excellence in Maths Teaching for Mastery Randomised Controlled Trial, Evaluation Report. University of Nottingham.

[21] Initially, we planned to stratify by both setting and group allocation in the efficacy trial (i.e., whether participants were in the full intervention or not). However, since none of the teachers participated in the efficacy trial, stratification was limited to setting only. The evaluation protocol has been updated to reflect this change.

[22] This additional analysis was agreed after publication of the evaluation protocol (version September 2024). The Mastering Maths programme was designed for students who previously achieved below a grade 4 in GCSE Maths and is expected to benefit all students regardless of their prior attainment. We are interested in exploring whether this is the case, or whether there is a differential impact for lower- or higher-attaining students.

in both scales are suitable for this student demographic, and that students understand what they are being asked. Students who take part in the cognitive interviews will receive a £30 voucher as a thank you for their participation. Results from the cognitive testing will inform whether any revisions are needed to the scales before they are rolled out during the main trial.

Participant selection

Setting and teacher eligibility

UoN will identify and recruit eligible FE colleges for this trial. All FE colleges in England with GCSE Maths resit students will be eligible. As FE colleges account for 70% of maths resit students[23], sixth form colleges will generally not be included in this trial. However, UoN will make ad-hoc judgements on the eligibility of sixth form colleges; those with large resit class cohorts (i.e., having at least 50 resit students) may be deemed as eligible.

Many FE colleges are large, meaning that the teaching of groups of GCSE Maths resit classes is separated. There can be minimal contact between groups of GCSE Maths resit classes, even if teachers are on the same campus. Discussion of teaching and teaching methods between groups can be rare. Therefore, for the purpose of this study, each of these groups is defined as a separate setting. Within one FE college, there can be several teachers from different settings eligible for the trial if they are geographically or structurally separate in terms of organisational set-up. Throughout the recruitment period, we will revisit the definition of 'setting' with UoN in order to capture any nuances in organisational set-ups.

UoN will recruit up to two GCSE Maths resit teachers from each setting. Settings will be encouraged to have two teachers participating in the trial but will still be accepted if they can only release one.

Teachers will not be eligible for the trial if they:

- took part in the full intervention arm of the Mastering Maths efficacy trial as either a lead teacher or intervention teacher;
- took part in the National Centre for Excellence in the Teaching of Mathematics (NCETM)'s Teaching for Mastery "Trailblazer" cohort during the 2023/24 academic year, or are taking part in "Cohort 1" of the programme during the 2023/24 academic years[24];
- are taking part in programmes involving substantial professional development during the 2024/25 academic year.

These eligibility criteria will ensure that both intervention and teaching-as-usual teachers are not exposed to similar programmes which may affect their teaching practice. This will enable us to estimate the programme's impact more accurately. In all settings, teacher(s) participating in the

[23] Smith, A. (2017). Smith review of post-16 mathematics: Report and letter. GOV.UK. https://www.gov.uk/government/publications/smith-review-of-post-16-maths-report-andgovernment-response;

[24] While we will not allow individual teachers to participate in both Mastering Maths and 'Teaching for Mastery', we will allow settings to have teachers participate in both during the 2024-2025 academic year. This is because at the setting level the two programmes are different, and the 'Train the Trainer' element of Teaching for Mastery will only begin after our evaluation.

trial and a senior leader will sign a Memorandum of Understanding (MoU), which indicates their commitment to the trial.

In the efficacy trial, the COVID-19 pandemic had a significant impact on attrition rates. To build upon this learning and manage attrition within the current trial, incentives will be provided for participation. Settings nominating one teacher will receive an incentive of £1,250 if allocated to the Mastering Maths intervention group (this is a payment of £250 to cover classes for each of the five lesson study meetings), in addition to receiving the Mastering Maths professional development. Those allocated to the teaching-as-usual group will receive an incentive of £1,000 as a thank you for participation and for facilitating data collection. Settings nominating two teachers will have one teacher allocated to the intervention group and one to the teaching-as-usual group. They will receive £1,250 for the teacher allocated to the intervention group to cover classes and the Mastering Maths professional development, and an additional £500 as a thank you for the participation of the teaching-as-usual teacher and the associated collection of student data. UoN will administer the incentives following endline data collection (likely to be October 2025), conditional on the completion of evaluation activities. Finally, to aid recruitment and retention, the settings whose teachers are allocated to the teaching-as-usual group will be granted priority access to the programme the following academic year.

Student eligibility

Eligible students will be those aged 16-19 who will resit their GCSE Maths exam in the summer of 2025 because they have not previously received a grade 4 or higher, and who are therefore enrolled in the resit classes of participating teachers as a condition of funding for their education.

Some students may have the opportunity to resit their GCSE Maths exam in November 2024. This is dependent on colleges' resit policy; some colleges may require all, or some, of their students to resit their exam in November 2024, rather than waiting until the following year. Students resitting the exam in November will receive their results in January 2025. This may affect their attendance at maths lessons in the interim; if students believe they have achieved a grade 4 or higher in the November resit, they may not attend the first few Mastering Maths or teaching-as-usual lessons. In this trial, we will include both those who do not take the November resit and those who do take it, but do not achieve a grade 4 or above and will subsequently be enrolled to resit their Maths GCSE in the summer of 2025. We will not include those who take the November resit exam and are no longer in the participating teacher's class in February because they have achieved a grade 4 or higher in the exam. This approach maximises the sample while capturing a real-world scenario appropriate for an effectiveness trial.

In November 2024, settings will inform students about the Mastering Maths study. Teachers will be asked to share an information sheet about the study with their GCSE Maths resit students and to withhold the randomisation allocation from them during the recruitment stage.[25] The information sheet will state that their participation is voluntary and that they have the right to withdraw throughout the evaluation, by contacting their teacher or the NatCen evaluation team.

[25] As students will be informed about the study after randomisation, teachers will be asked to withhold the randomisation allocation from their students during the recruitment stage. This will prevent students on making their decision about participation based on the results of randomisation allocation, thus minimising the risk of introducing bias in our study.

Withdrawn students will not be part of the evaluation but will continue to attend the Mastering Maths lessons, if their teacher is allocated to the intervention group.

Data collection for Impact Evaluation

We will collect student- and setting-level data directly from settings and teacher-level data from UoN. Settings will provide contact details of teachers to UoN during trial recruitment. UoN will then provide NatCen with teacher-level data (names, email addresses, name of setting, phone number, postal address) prior to randomisation, alongside the list of settings. Lead teachers will also collate attendance data on whether teachers attended training. In addition, teachers will self-report to lead teachers which and how many of the Mastering Maths lessons were taught to each class. Lead teachers will then report this information to UoN, who will pass it to the evaluation team.

In November 2024, teachers will provide student information through NatCen's secure upload platform, including a list of full names, post codes, dates of birth and Unique Learner Numbers (ULNs). We will use these details to obtain information from the NPD on students' prior attainment in maths at both KS2 and Key Stage 4 (KS4), and their prior FSM eligibility.

In February 2025, teachers will be asked to provide an updated student list that excludes those students who will no longer be in their class because they achieved a grade 4 or higher in the November resit exam. We will not classify these students as attrition since they do not meet our eligibility criterion of needing to resit the Maths GCSE exam in the summer of 2025. All other students will be included in the final sample, including students who received a grade 3 or below on the November resit exam and those who did not take the November resit. Only students who should be in the final analytical sample, but aren't, will be classified as attrition. Student attrition will thus be defined as:

$$(N\ in\ the\ inital\ student\ list - N\ passing\ the\ November\ resit) \\ - N\ in\ the\ final\ analytical\ sample$$

In March 2025, teachers will be asked to administer a paper-based student survey during a maths lesson. We will collect data on students' self-confidence and self-efficacy in maths using this survey. Please see the endline surveys section under the IPE for more information.

In the summer and autumn of 2025, after the results of the spring GCSE exams are made available, we will receive the students' GCSE Maths raw scores and the raw score conversion tables for the exam board. We will liaise with the data managers in colleges and teachers as appropriate, to facilitate this process.

Randomisation

After UoN has completed recruitment in early September 2024, a NatCen analyst will randomise teachers into those being part of the Mastering Maths programme or those continuing teaching-

as-usual. Teachers recruited for the trial will be stratified by FE setting.[26] This stratification ensures that teachers within the same setting (or within the group of teachers who are the sole participants from their setting) are allocated evenly between intervention and teaching-as-usual groups). In settings with two teachers, we will randomly assign one teacher to the Mastering Maths intervention group and the other to the teaching-as-usual group. We will then group all other teachers (e.g., from settings with only one teacher) into a single stratum and randomly assign half to the Mastering Maths intervention group and the other half to the teaching-as-usual group. While this method ensures a balanced intervention allocation among all participating teachers, it does not guarantee balance within strata, as the odd number of teachers within a stratum cannot be equally divided into two groups.

We had originally planned to randomise at the setting-level but changed to teacher-level randomisation to help meet the recruitment targets. A key consideration here was the risk of spillover, namely from teachers from the intervention and teaching-as-usual groups being in the same setting. Having considered this, we do not anticipate serious spillover issues with the teacher-level randomisation approach. The efficacy trial showed that participation in lesson study sessions is the main driver of the programme's impact. The lesson study sessions provide teachers with collaborative learning with other teachers across a cluster of settings, and ensure teachers adopt the Mastering Maths teaching approach rather than just using the provided materials. During the 2024/25 academic year, teachers randomly allocated to the teaching-as-usual group will not take part in the professional development or lesson study sessions, which minimises the risk of potential spillover between teachers. Therefore, while the programme design minimises the risk of potential spillover due to teacher-level randomisation, this randomisation approach allows us to have adequate statistical power with smaller sample size. The IPE will, nonetheless, explore any spillover effects that may take place. Please see the IPE section for more information.

We will randomise teachers in September 2024. During randomisation, analysts will be blinded to teacher and setting identity; identifiers will only be merged with group allocation data after randomisation. The randomisation process will be conducted in Stata using the ***randtreat*** command. Both do and log files will be used to record the randomisation process.

Outcome measures

Baseline measures

For this evaluation, eligible students will be those aged 16-19 who will resit their GCSE Maths exam for having previously not attained a grade 4 or higher. Our target population could only receive grades 1, 2 and 3 in their KS4 GCSE maths exam, and thus the measure would have lower variation and lower explanatory power to predict their maths attainment. To address this, we will use KS2 maths attainment as the baseline measure for all outcome analyses, including attitudes towards maths (RQ7-8).

[26] Initially, we planned to stratify by both setting and group allocation in the efficacy trial (i.e., whether participants were in the full intervention or not). However, since none of the teachers participated in the efficacy trial, stratification was limited to setting only. The evaluation protocol has been updated to reflect this change.

At the end of KS2, pupils undertake standardised tests in maths based on the national curriculum, which include three papers: one arithmetic and two reasoning tests. The arithmetic test focuses on calculations and lasts 30 minutes, while each reasoning test assesses problem-solving skills and understanding of mathematical concepts and lasts 40 minutes. These KS2 National Curriculum Tests in maths are externally marked, and the raw and scaled scores are returned to schools and made accessible to researchers in the NPD. We will use the raw score ('KS2_MATMRK' variable from the NPD) as the baseline measure, which ranges from 0 to 110.[27]

Primary outcomes

Our population of interest is students resitting their GCSE Maths as a condition of funding for their education. The primary outcome of interest is their performance in GCSE Maths exams. GCSE Maths exams are administered by different exam boards in different settings, at the end of the academic year, in May and June. The exams comprise multiple papers that test a range of mathematical skills, including number, algebra, ratio and proportion, geometry and measure, probability and statistics. There are three papers: one non-calculator paper and two calculator papers, each lasting 90 minutes. The exams are divided into foundation and higher tiers, where the foundation tier covers grades 1-5 and the higher tier covers grades 4-9. The decision on which tier to take is largely influenced by teachers as well as students' past performance.

GCSE Maths scores are converted to grades on a 1-9 scale, with 9 being the highest. Grade 4 is a 'standard pass' and grade 5 is a 'strong pass'. These scores are standardised to maintain equivalence in grades regardless of the exam board, so that all students are assessed consistently and fairly. We expect most students in our target population to sit the foundation paper only. This limits the grade scale from 1-9 to 1-5 and prevents us from distinguishing between different levels of student performance. This lack of differentiation reduces the sensitivity of the measure and the power of analysis. For this reason, we will instead use GCSE Maths raw scores, collected directly from colleges. These range from 0 up to around 80 or 100 per paper, depending on the exam board or year of assessment, and are considered a continuous measure. The granularity of the raw scores allows for more detailed differentiation between students and enhances the sensitivity of the measure and the power of our analysis. This is in contrast to the categorial data of GCSE Maths grades that may not detect finer differences like raw scores can.

As different exam boards use different scales, this complicates comparison of the raw scores across the four exam boards.[28] To address this, we will first group students by exam board. Within each group, we will standardise the GCSE raw scores to have a mean of 0 and standard deviation of 1. After standardisation, we will combine these standardised scores from all exam boards to create a single primary outcome measure.

[27] We initially planned to use the Maths scaled score (KS2_MATHSCORE_noSpeccon variable, 0-120) due to uncertainties surrounding the availability of raw maths attainment scores in the NPD data. However, after communicating with the Department for Education, we confirmed that the raw scores, labelled as 'marks', are available in the NPD. We then choose using the raw score ('KS2_MATMRK' variable) as a baseline measure, in line with the EEF's statistical guidance (2022).

[28] The exam boards are Assessment and Qualifications Alliance (AQA), Pearson Edexcel, Oxford Cambridge Recognition (OCR) and WJEC.

Secondary outcomes

We will analyse two secondary outcomes that are related to GCSE Maths results: the probability of students a) achieving a grade 4 or higher (RQ2) and b) moving up a grade (RQ3). For these outcomes, we will first convert the GCSE Maths raw scores into a grade based on the relevant conversion tables provided by exam boards. For RQ2, we will then create a binary measure which will take a value of 1 if a student achieves a grade 4 or higher, and 0 if otherwise. Similarly, for RQ3, we will create a binary measure taking a value of 1 if a student receives a higher grade than in their previous GCSE Maths exam and 0 if otherwise.

We will also estimate whether the impact of Mastering Maths differs for students from disadvantaged backgrounds as measured by prior FSM status (RQ4) and the Index of Multiple Deprivation (IMD) (RQ5) [29]. For RQ4, we will include an interaction term on students' prior FSM status using "EVERFSM_6_P_[term][yy]" from the NPD, which indicates if a student has been recorded as eligible for FSM at any point in the last six years. For RQ5, we will include an interaction term for IMD status, a binary indicator identifying students whose home postcodes fall within the 27% most deprived areas, based on the IMD 2019 and the list of eligible postcodes published annually by the Education and Skills Funding Agency (ESFA).[30] The IMD is the official measure of relative deprivation in England, ranking areas from the most deprived (rank 1) to the least deprived, and categorising them into deciles from 1 (most deprived) to 10 (least deprived).

We will also estimate whether the impact of Mastering Maths differs for students with lower prior attainment (RQ6).[31] The Mastering Maths programme was designed for students who previously achieved below a grade 4 in GCSE Maths and is expected to benefit all students regardless of their prior attainment. We aim to explore whether this expectation holds or whether there is a differential impact for lower- versus higher-attaining students. Consistent with RQ4 and RQ5, we will include an interaction term for low prior attainment, defined as having previously received a grade 2 or below in GCSE Maths, derived from the NPD.

The programme logic model indicates that a short-term outcome of Mastering Maths is an improvement in student's confidence and self-efficacy in maths. To measure this (RQ7), self-confidence will be captured using the Self-Confidence subscale of the Attitudes Towards Maths Inventory (ATMI), which has been widely used elsewhere[32,33,34,35]. For self-efficacy (RQ8), we will use a questionnaire (self-confidence subscale) developed by the University of Manchester. We will provide more detailed information on these in the Statistical Analysis Plan (SAP).

[29] English indices of deprivation 2019. Ministry of Housing. Available at https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019

[30] 2024 to 2025 uplift factors by postcode and LLSOA. ESFA. Available at https://www.gov.uk/government/publications/uplift-factors-and-postcode-files

[31] This additional analysis was agreed during the peer review of the SAP after publication of the first version of the evaluation protocol.

[32] Tapia and Marsh, 2000. https://eric.ed.gov/?id=ED449045

[33] Tapia and Marsh, 2004. http://www.rapidintellect.com/AEQweb/cho25344l.htm

[34] Hanley et al, 2021. https://d2tic4wvo1iusb.cloudfront.net/production/documents/pages/projects/5Rs_post-16_GCSE_Resit_maths_COVID_clean.pdf?v=1696581772

[35] Dysart et al, 2023. https://d2tic4wvo1iusb.cloudfront.net/documents/projects/V1_5Rs-EEF_trial_protocol_2022.docx_CLEANMarch2023.pdf?v=1682079782

Sample size

Table 2: Power calculations

Assumptions		Without Attrition All students	Without Attrition FSM eligible students (36%)	With Attrition: Teacher: 15% Student: 25% All students	With Attrition: Teacher: 25% Student: 55% All students	With Attrition: Teacher: 15% Student: 25% FSM eligible students (36%)	With Attrition: Teacher: 25% Student: 55% FSM eligible students (36%)
Minimum Detectable Effect Size (MDES)		0.188	0.206	0.206	0.227	0.227	0.259
Pre-test / post-test correlations	Level 1 (student)	0.25					
	Level 2 (teacher)	0.23					
Intra-cluster correlations (ICCs)	Level 2 (teacher)	0.14					
Alpha		0.05					
Power		0.80					
One-sided or two-sided		2					
Number of Level 2 covariates[36]		105	105	90	80	90	80
Average cluster size		50	18	44	30	15.75	10.7
Number of teachers	Mastering Maths	70	70	60	53	60	53
	Teaching-as-usual	70	70	60	53	60	53
	Total	140	140	120	106	120	106
Number of students	Mastering Maths	3500	1260	2625	1575	945	567
	Teaching-as-usual	3500	1260	2625	1575	945	567
	Total	7000	2520	5250	3150	1890	1134

[36] The number of Level 2 covariates are based on our estimate of the number of settings with one and two teachers, for which we assumed an even split (half of settings will have one teacher and the other half two teachers). The power calculations will be updated in the SAP and these will reflect the accurate number of Level 2 covariates.

Table 2 shows the estimated Minimum Detectable Effect Size (MDES), with and without accounting for attrition at both the teacher and student level, for the primary outcome in the trial. Our power calculations are informed by the Mastering Maths efficacy trial.[37] We assume a Type I error rate of 0.05 and a Type II error rate of 0.20 (i.e., 80% power). We use estimates found in the efficacy trial for the correlation between baseline (KS2 Maths score) and endline (GCSE Maths score) attainment at the student level (0.25) and at the teacher level (0.23), and for the intra-cluster correlation (0.14).[38] We assume 50 eligible students per teacher on average. We have conducted the power calculations using PowerUp! (Dong & Maynard, 2013).

Attrition is a risk for any study and a key consideration for this trial. The efficacy trial was heavily impacted by the COVID-19 pandemic (especially by national and regional lockdowns, frequent staff and/or student absences, and teachers' additional workload). The student attrition level from recruitment to analysis was around 55% and was similar between those receiving the Mastering Maths intervention and those continuing teaching as usual. The setting attrition level was around 24%, with a lower level of attrition for those assigned to the teaching-as-usual group compared to the full and partial intervention groups. It is also important to note that the attendance of students in the resit exam (i.e., non-response) may exacerbate the student attrition level. Therefore, we expect significant attrition from baseline recruitment numbers at both the student and teacher levels, though potentially lower than in the efficacy trial given the COVID-19 context.[39]

Table 2 is based on the assumption that around 120 teachers will remain in the trial at endline (representing roughly 15% attrition), with an average of 44 students per teacher (accounting for 25% attrition at the student level). Under these assumptions, we would achieve an MDES of about 0.206. If the attrition levels were similar to the efficacy trial (i.e., 25% at the teacher level and 55% at the student level), the trial would achieve an MDES of approximately 0.227.

We have also conducted power calculations to estimate the MDES for subgroup analysis according to prior FSM eligibility. Based on the efficacy trial, we assume that 36% of students are eligible for FSM before further education. This indicates the trial would be able to achieve an MDES of around 0.227.

The efficacy trial had sufficient power to detect an effect size of 0.24 at the analysis stage.[40]. Table 3 below shows the sample size requirement at the analysis stage to detect the same effect size. As such, to detect an MDES of 0.24, our minimum acceptable sample size for this trial is 88 teachers, with an average number of 44 students per teacher.

[37] Wake, G., et al. 2023. Centres for Excellence in Maths Teaching for Mastery Randomised Controlled Trial, Evaluation Report. University of Nottingham.

[38] In the efficacy trial, there was one teacher per setting. We, therefore, assumed that setting level estimations from the efficacy trial be applicable to the teacher level estimations for this trial.

[39] As some of this high attrition may have been related to the COVID-19 pandemic, we anticipate lower rates of attrition during this trial.

[40] This effect size represents our 'maximum acceptable sample size' scenario, where we would still proceed with the analysis, despite the EEF's padlock rating system.

Table 3: Power calculations continued

Assumptions		Recruitment Stage: All students	Analysis Stage: All students
Minimum Detectable Effect Size (MDES)		0.221	0.24
Pre-test / post-test correlations	Level 1 (student)	0.25	
	Level 2 (teacher)	0.23	
Intra-cluster correlations (ICCs)	Level 2 (teacher)	0.14	
Alpha		0.05	
Power		0.80	
One-sided or two-sided		2	
Number of Level 2 covariates		70	60
Average cluster size		50	44
Number of teachers	Mastering Maths	51	44
	Teaching-as-usual	51	44
	Total	102	88
Number of students	Mastering Maths	2550	1936
	Teaching-as-usual	2550	1936
	Total	5100	3872

Statistical analysis

Primary analysis

We will use the Intention to Treat (ITT) approach to estimate the impact of Mastering Maths on the primary outcome measure (RQ1). We will use a two-level linear regression model with students (Level 1) being clustered within teachers (Level 2). The dependent variable will be students' GCSE

Maths standardised raw score. The independent variables will be a binary indicator of intervention allocation, KS2 maths attainment raw score (as a baseline measure), exam board fixed effects[41] and the randomisation strata (setting). Teacher-level random effects will also be included in the model by allowing the intercept to vary randomly across teachers. Following the EEF statistical analysis guidance[42], the impact of the programme will be expressed as a standardised effect size using Hedges' *g* with 95% confidence intervals.

Secondary analysis

We will also estimate the impact of Mastering Maths on secondary outcomes using an ITT approach. We will use a two-level logistic regression model with students (Level 1) being clustered within teachers (Level 2) to estimate the effect of the Mastering Maths programme on students' probability of achieving a grade 4 or higher (RQ2), and their probability of moving up a grade (RQ3), on the GCSE Maths exam. The analytical approach for both research questions will be analogous to the primary outcome estimation.

We will estimate the impact of Mastering Maths on students' self-confidence (RQ7)and self-efficacy in maths (RQ8) following the primary analysis model. As in the primary analysis model, we will use KS2 maths attainment as a baseline measure rather than a baseline measure of students' attitudes towards maths. This is because the relevant surveys will only be administered at endline, to reduce burden on teachers and students. Nonetheless we believe students' prior attainment in maths is correlated with their attitudes towards the subject and can thus explain some of the variance in the model.

Sub-group analysis

To estimate if the impact differs for students from disadvantaged backgrounds (RQ4 and RQ5) and by prior attainment (RQ6), we will conduct sub-group analysis. Each analysis will follow the primary model and include an interaction term:

- RQ4: Prior FSM status, using the "EVERFSM_6_P_[term][yy]" variable from the NPD.
- RQ5: IMD status, identifying students whose home postcodes fall within the 27% most deprived areas, based on the list of eligible postcodes published annually by the ESFA.
- RQ6: Low prior attainment, defined as having previously received a grade 2 or below in GCSE Maths, derived from the NPD.

Analysis in the presence of non-compliance

We will also explore programme effects for teachers allocated to the intervention group who implemented the intervention as intended, based on compliance criteria. Compliance is defined as the fulfilment of a set of minimum criteria which determine whether a teacher has run the programme as intended. This will be a binary measure, indicating whether a teacher was

[41] Exam board fixed effects will be included as a covariate to account for differences between the four exam boards delivering the exams.
[42] EEF, 2022. EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf (d2tic4wvo1iusb.cloudfront.net)

compliant or not. We will define the compliance measure at the teacher level but conduct the compliance analysis at the student level.

We agreed the compliance criteria together with the delivery team and the EEF. For a teacher to be deemed compliant, they need to fulfil all three activities[43]: teacher professional development, lesson study, and teaching of Mastering Maths lessons. For full compliance, teachers first need to attend both days of professional development. In instances where teachers cannot attend teacher training on those dates or times, attending mop-up training sessions will also count as full compliance. Teachers also need to attend at least four out of the five lesson study group sessions and teach five Mastering Maths lessons to all their GCSE resit classes. Lead teachers will monitor attendance during professional development days and lesson study groups. They will also monitor whether teachers teach all five Mastering Maths lessons that their lesson study group choose, and if they taught any additional lessons. Lead teachers will securely share this information with the UoN, who will pass this on to the evaluation team. The compliance criteria are summarised below in Table 4.

Table 4: Compliance criteria table

COMPLIANCE CRITERION	DATA SOURCE	COMPLIANCE INDICATOR
Attendance in professional development	Attendance register completed by lead teacher	Teacher attends both days of professional development (this can include a mop-up day if they miss their original session)
Attendance at lesson study sessions	Attendance register completed by lead teacher	Teacher attends at least four out of five lesson study sessions
Teaching of Mastering Maths lessons	Compliance register completed by lead teacher	Teacher teaches five Mastering Maths lessons

In cases of *teacher* attrition, defining compliance for students whose teacher drops out will be done on a case-by-case basis.

We will follow the EEF's statistical analysis guidance to address non-compliance.[44] A Complier Average Causal Effect (CACE) analysis will be conducted using an Instrumental Variables (IV) approach[45]. We will use a two-stage least squares estimation approach to recover the intervention effect for compliers. The first step will estimate the compliance rate, i.e., whether being assigned to Mastering Maths influences whether teachers comply. The second step will substitute the intervention indicator with the estimated compliance rate from the first step to predict the outcome. The results of this model will answer the research question: 'To what extent

[43] Refer to the 'Intervention' section for more information on the Mastering Maths programme.
[44] EEF, 2022. EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf (d2tic4wvo1iusb.cloudfront.net)
[45] Imbens and Angrist, 1994. https://business.baylor.edu/scott_cunningham/teaching/imbens--angrist---late-1994.pdf

does compliance with the Mastering Maths programme lead to improved GCSE Maths outcomes?'. This model will be estimated for the primary outcome measure only.

Missing data analysis

We will follow EEF statistical guidance to handle missing data and follow the definition of student attrition explained in the Data Collection for Impact Evaluation section above. For less than 5% overall missingness from randomisation to final analysis, a complete-case analysis will be employed. In the event that more than 5% of data is missing overall from baseline to final analysis, our approach will depend on the pattern of missing data. If the pattern is unrelated to observable and/or unobservable variables, then we will assume data is Missing Completely at Random (MCAR), and we will continue with complete case analysis. If data is missing in a way that is correlated with observable variables, then primary analysis will be re-estimated through Multiple Imputation using Chained Equations (MICE). More information on the missing data calculations will be provided in the SAP.

Sensitivity analysis

Students resitting and receiving grade 3 or below on November resit exams are likely to have lower attendance rates between November 2024 and January 2025 compared to those who only plan to resit in the summer. This is because the results of the November resit exams will be available in January 2025, and students may disengage from their maths learning while waiting for their results. In an ideal scenario, we would have student attendance data to account for dosage. However, this would be overly burdensome for teachers to collect and share. Instead, we will collect information on whether students attempted the November 2024 GCSE Maths resit, using student enumeration. This will serve as a proxy for attendance, under the assumption that students who sat the November exam had lower lesson attendance during this period.

Sensitivity analyses will assess whether November resit exams influence the results. We will replicate the primary outcome analysis by including a binary indicator for November resit participation to account for potential differences in attendance between students who resit and those who do not.[46] Additionally, we will repeat both the primary and CACE analyses after excluding November resit students, who are more likely to disengage. If these analyses produce results consistent with the original models, this would suggest that the intervention effect is robust to November resit participation; if they differ, it may indicate that attendance patterns affect the intervention's impact and should be considered in future trials.

Finally, in line with the EEF's statistical analysis guidance (EEF, 2022), we will also estimate the primary outcome model including only prior attainment and the stratification variable. This analysis will be undertaken solely to support the EEF's synthesis work, and the results will be presented in an appendix. They will not be compared with those from the primary outcome analysis.

[46] A binary variable indicating whether a student took the November resit exam was originally included to the primary analysis to account for variations in attendance. However, due to a high level of missing data for this variable, we decided to conduct a separate sensitivity analysis instead.

Mediation analysis

Mediation analysis is used to explore mechanisms by which an intervention affects the outcome of interest. For Mastering Maths, we propose two mechanisms by which the intervention affects the GCSE resit outcomes of interest: improving students' self-confidence and self-efficacy in maths. This is captured in the programme logic model.

We will conduct an exploratory mediation analysis to understand whether the effect of Mastering Maths on the primary outcome is partially or totally mediated by changes in students' self-confidence and self-efficacy in maths. Although self-confidence and self-efficacy are treated separately in the secondary outcome analysis (as they come from different measures), for mediation analysis we will combine them into a single mediator representing attitudes towards maths. This approach is justified because self-confidence and self-efficacy are likely to have a bidirectional causal relationship. That is, self-efficacy on specific maths tasks can influence general confidence, while confidence can enhance self-efficacy, making separate mediation estimates inappropriate.[47] To operationalise this combined mediator, we will sum z-score standardised self-confidence and self-efficacy scores, given the different scales used for the two measures. Mediation analysis will be outlined in more detail together with its assumed causal model in the SAP.

Estimation of effect sizes

We will use the effect sizes (ES), as adapted from Hedges (2007), for cluster-randomised trials (as per EEF statistical guidance (2022)):

$$ES = \frac{(\overline{Y_T} - \overline{Y_C})_{adjusted}}{\sqrt{\sigma_u^2 + \sigma_e^2}}$$

Where $(\overline{Y_T} - \overline{Y_C})_{adjusted}$ is the difference in means between the intervention and teaching-as-usual group adjusted for baseline characteristics and stratification variables. $\sqrt{\sigma_u^2 + \sigma_e^2}$ is an estimate of the population standard deviation. σ_u^2 is the variance of teacher level intercept and σ_e^2 is variance of residuals.

For RQ 1, 4 and 5, we will use each group's adjusted mean and variance from the primary and secondary outcome model to calculate the effect size. The variance will be the total variance (across both students and teachers), without any covariates, as emerging from a 'null' or 'empty' multi-level model with no predictors. We will report a 95% confidence interval that accounts for the clustering.

[47] As detailed by Imai and Yamamoto (2013), standard causal mediation relies on the ignorability assumption, which requires no post-treatment confounding between the mediator and the outcome, whether observed or unobserved. When multiple mediators are causally independent, separate mediation effects can be estimated for each mediator. However, self-confidence and self-efficacy are likely to exhibit a bidirectional causal relationship, making separate mediation estimates inappropriate. For this reason, they are treated as two closely related constructs or alternative measures of a common underlying construct: attitudes towards maths.

Following EEF statistical guidance (2022), in the case where an outcome measure is defined as a binary variable (i.e., RQ 2 and 3), we will transform ES measures into ES measures comparable to Hedges' g. More details on the effect size calculations will be included in the SAP.

Implementation and process evaluation (IPE) design

Research questions

The IPE of Mastering Maths will address the research questions and dimensions set out in Table 5. Our IPE dimensions of interest are informed by EEF's framework for implementation and process evaluation.[48] The design explores to what extent the programme is implemented as described in the logic model and whether it is delivered well. Complementing our impact evaluation design, it collects evidence on the intervention and teaching-as-usual groups and the differences between disadvantaged students and their more advantaged peers. In addition, our IPE inspects whether the short- and long-term outcomes take place as described in the logic model and what the causal relationships are between outcomes and moderators/contextual factors. Lastly, we seek to understand the cost of Mastering Maths.

Table 5: IPE research questions and dimensions

Research question	Dimension
1. To what extent is Mastering Maths delivered as intended, particularly the new 'Train the Trainer' model?	Fidelity
2. What adaptations are made to the intended model, and why?	Adaptation
3. How many training sessions, lesson study groups, and maths lessons are delivered?	Dosage
4. How well are the training sessions, lesson study groups and maths lessons delivered?	Quality
5. What is the rate and scope of participation in training sessions, lesson study groups and maths lessons? a. Is reach different for disadvantaged students?	Reach
6. How well do teachers, lead teachers and students engage with Mastering Maths? a. What are the barriers and facilitators to engagement, particularly for disadvantaged students?	Responsiveness
7. How is Mastering Maths different to usual practice for GCSE Maths resit classes? a. How are Mastering Maths lessons different to usual GCSE Maths resit teaching? b. Do intervention group teachers take part in other similar professional development programmes to Mastering Maths?	Differentiation
8. How is Mastering Maths different to the teaching-as-usual condition? a. What is taking place instead of Mastering Maths (teaching-as-usual)? b. Do teaching-as-usual teachers take part in other similar professional development programmes to Mastering Maths?	Monitoring the control
9. What impacts do teachers, lead teachers and students perceive to result from Mastering Maths?	Perceived impacts

[48] Humphrey, N. et al. 2019. Implementation and process evaluation (IPE) for interventions in education settings: an introductory handbook.

a. Do perceived impacts differ between disadvantaged and more advantaged students?	
10. Which contextual/moderating factors influence programme delivery/impacts?	Context/moderators
11. What is the cost of Mastering Maths?	Costs

Research methods

NatCen will take a mixed-method approach to IPE data collection. The IPE design has been carefully thought through to reduce burden on busy teachers who will also be required to complete tasks for the impact evaluation. For example, we have not included baseline teacher surveys as these would need to take place at the same time that teachers are required to attend professional development days and enumerate students for the trial in the autumn of 2024. Instead, both teachers and students will complete one paper-based survey towards the end of the trial. The activities are sequenced so that we conduct research activities evenly throughout the academic year. This is both to reduce burden and to allow us to collect data at different points during intervention delivery: at baseline (training observations, baseline lead teacher interview), during intervention delivery (case studies) and endline (surveys, endline teacher and delivery team interviews). The activities will collect data against the IPE dimensions and research questions; they are not designed to be iterative.

Method	IPE dimension										
	Fidelity	Adaptation	Dosage	Quality	Reach	Responsiveness	Differentiation	Monitoring control	Perceived impacts	Costs	Context/moderatos
Training observations											
Baseline lead teacher interview											
Endline lead teacher interview											
Observation of teacher's lesson (intervention case study)											
Observation of teacher's lesson study group (intervention case study)											
Paired or 1:1 student interviews (intervention case study)											
Teacher interview (intervention case study)											
Observation of teacher's lesson (teaching-as-usual case study)											
Paired or 1:1 student interview (teaching-as-usual case study)											
Teacher interview (teaching-as-usual case study)											

Endline surveys (intervention)											
Endline surveys (teaching-as-usual)											
Delivery team data collection											

Figure 2 below maps our research methods against the key IPE dimensions.[49]

Figure 2: IPE methods and dimensions

Method	IPE dimension										
	Fidelity	Adaptation	Dosage	Quality	Reach	Responsiveness	Differentiation	Monitoring control	Perceived impacts	Costs	Context/moderatos
Training observations											
Baseline lead teacher interview											
Endline lead teacher interview											
Observation of teacher's lesson (intervention case study)											
Observation of teacher's lesson study group (intervention case study)											
Paired or 1:1 student interviews (intervention case study)											
Teacher interview (intervention case study)											
Observation of teacher's lesson (teaching-as-usual case study)											
Paired or 1:1 student interview (teaching-as-usual case study)											
Teacher interview (teaching-as-usual case study)											
Endline surveys (intervention)											
Endline surveys (teaching-as-usual)											
Delivery team data collection											

We will observe two out of the five days of face-to-face training delivered to lead teachers over June and September 2024. Once the training content has been confirmed, we will liaise with the delivery team at UoN to find the most suitable days for us to observe training. We will also

[49] Changes to which IPE dimensions were captured by each IPE activity have changed since the first draft of the protocol. The dimensions captured during case studies were altered to better suit the data we were able to capture (e.g. Quality was removed from student interviews, as students cannot comment on Quality) and the dimension of Costs was removed from the endline lead teacher interview as they were felt to be best captured through a cost pro-forma, completed by the delivery team.

observe one regional cluster of the two face-to-face professional development days provided by lead teachers to intervention teachers in autumn 2024; similarly, the specific days will be agreed with UoN. Lastly, we will review all documentation that will be shared with teachers (e.g., handbook of key principles, example lessons and video clips). Our observation protocols will collect data on IPE dimensions of **fidelity**, **adaptation**, **quality**, and **responsiveness**.

Lead teacher interviews

We will interview six of the 15 lead teachers. The lead teachers will consist of individuals recruited because they have taken part in the previous efficacy trial, have received training in a relevant area or who have experience of delivering professional development. We will aim to interview three individuals who took part in the previous efficacy trial, and three individuals with other types of backgrounds. These six individuals will be interviewed twice, first in the autumn of 2024 (after the second set of lead teacher training) and again at endline in the summer of 2025 (after they have run the lesson study groups). We will receive lead teachers' contact details from the delivery team and will contact individuals directly to invite them to take part.

The interviews will be up to 60 minutes in length and will take place online or by telephone. We will design semi-structured topic guides that will collect evidence on **fidelity, adaptation**, **dosage**, **reach, responsiveness**, **perceived impacts,** and **context/moderators**. The topic guides will be structured around the research questions outlined above in Table 5.

During project set-up, we identified two elements of the lead teachers' roles as needing particular attention in the IPE:

- Does the training lead to lead teachers understanding key elements of the programme?
- Does the training ensure that lead teachers feel capable of running the teacher training and lesson study groups?

Case studies

We will conduct eight case studies in February 2025 while intervention delivery is ongoing. The trial consists of two types of settings: those with two teachers and those with one teacher. After publishing the original version of the evaluation protocol, it was agreed that we will aim to recruit settings with two teachers taking part in the trial for all case studies so that we can explore whether there has been any contamination between intervention and control teachers. This is outlined in Table 6 below. In each of the settings, we will focus on a different teacher. In four settings we will focus on a teacher randomised to the intervention group, whereas in the other four settings, we will focus on a teaching-as-usual teacher.

Table 6: IPE case study numbers

Setting type	Number of settings selected for a case study	Number of teachers per settings
Intervention and teaching-as-usual	8	1

Intervention group

Intervention case studies will focus on teachers randomised to the intervention group. We will carry out a single two-day case study visit that includes an observation of the teacher hosting a lesson study group cluster meeting on one day, followed by a teacher interview, student interview, and an observation of them teaching their class without their lesson study group peers present on the other day.

The case study design will allow us to see how an individual teacher's experience of the regional professional development days and small-group lesson study translates to their application of Mastering Maths approaches in the classroom, and consequently the experiences and perceived outcomes of their students. We will also enquire about potential spillover between intervention and teaching-as-usual group teachers within a setting. The teacher is the key actor in the Mastering Maths intervention, either taking part in or providing most of the activities, and so they will be the unit of analysis.

We will organise the case studies with the help of lead teachers, as we need to know which teachers are hosting the lesson study groups in February 2025 when we plan to conduct case study visits. After receiving a list of names and contact details from lead teachers, we will contact teachers in the winter of 2024/25. We will sample teachers based on geography, in order to have a spread of geographic clusters used for lesson study sessions. We will give teachers who agree to help us with case studies £200 e-vouchers as a thank-you for their time. We will ask teachers to select two students to take part in a research interview ahead of our visit. If these students are absent on the day, or this has not been possible, we can ask students to take part on the day. We will give students who take part in interviews a £30 e-voucher as a thank-you for their participation.

We will organise case study visits on a day when the teacher is hosting a lesson study group at their setting. Case studies will be in person and will include the research activities below. The IPE dimension that each activity addresses is detailed in bold.

- Observation of teacher's Mastering Maths lesson – **fidelity, adaptation, quality, reach, responsiveness**
- Observation of teacher's lesson study group – **fidelity, adaptation, quality, reach, responsiveness**
- Student interviews (no longer than 30 minutes; with students being given the option of doing these individually or in pairs, to facilitate engagement but to avoid the potentially challenging group dynamics of focus groups; an approach successfully used in the 5Rs

evaluation[50]) – **fidelity, reach, responsiveness, perceived impacts, context/moderators**[51]

- Teacher interview (no longer than 50 minutes) – **fidelity, adaptation, reach, responsiveness, differentiation, perceived impacts, context/moderators**

We will design structured observation protocols and semi-structured topic guides for case studies, which will be designed around the research questions and IPE dimensions.

During project set-up, we identified two elements of the teachers' roles as needing particular attention in the IPE, and which will be addressed through case study interviews with teachers:

- Does the teacher professional development and the lesson study sessions, provided by lead teachers, lead to increased understanding of Mastering Maths principles of teaching and lesson study?

- Do the research questions used in lesson study sessions help to focus the teachers' observations of lessons and understanding of Mastering Maths?

During the set-up stage, we also agreed with the EEF and delivery team to use the case studies to collect evidence on college policies regarding the November resit exam. This is because the college policy (i.e., whether students are automatically enrolled or not, if students are charged for sitting the exam) and students' engagement with the exam in response (i.e., if they sat the exam, why or why not) is likely to affect the students' engagement with the intervention. For example, if a student sat the November resit exam and was hopeful that they had achieved a grade 4, they may stop attending maths lessons – only to find out later that they did not pass the exam. That is, by the time they receive their results for the resit exam, students might have already missed the first couple of Mastering Maths lessons, affecting the **reach** of the programme. This situation might be less likely to occur if, for example, the college's policy is to charge students for the additional exam, therefore discouraging a wider group of students from attempting the resit. For these reasons, the case study interviews with teachers and students will explore their college's general policy towards the November resit exam, how students engaged with the exam and lessons in response, and if this in turn affected reach.

The question on college policies for the November resit exam also relates to the IPE dimension of **context/moderators**. In interviews with teachers, we will ask about other contextual factors that may influence intervention delivery or impacts, such as senior leadership buy-in, student attendance, staff turnover and proportions of FSM students in the setting.

Lastly, the student interviews will collect information about **student disadvantage status**. This is one of the contextual/moderating factors in the logic model: that socially disadvantaged

[50] EEF, The 5Rs approach to approach to GCSE Maths resits (21/22 and 22/23 trial), Association of Colleges https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/the-5rs-approach-to-gcse-maths-resits-accelerator-fund

[51] In the original version of the protocol, all student interviews were planned to be in pairs, without giving students the option. We revised this approach to accommodate student preferences and offer more flexibility.

students will benefit from the programme more compared to their more advantaged peers, as was found in the efficacy trial. At the end of interviews[52], we will ask students to self-disclose whether they have ever been in receipt of FSM. For ethical reasons, because the students may be paired with their peer, we will explain to the students that they do not have to say yes if they do not wish to do so. During the analysis stage, we will then gauge whether the students' answers differ based on their self-disclosed FSM status.

Students self-disclosing their FSM status only at the end of the interview means that we will not be able to sample students to take part in interviews on the basis of their disadvantage status. This is because teachers who will help us organise the research activities will not themselves know students' FSM status from school as this information is not passed on from secondary schools to FE settings. For ethical reasons, we will not ask teachers to request all their students to self-disclose this information for the purpose of our sampling framework. Lastly, it is not possible to obtain information about students' past FSM statuses from the NPD and use this to sample students for interviews. This is because information from the NPD used for the impact evaluation analysis will only be available to the impact evaluation team on the Office for National Statistics' Secure Research Service (SRS) and cannot be transported out of the SRS to be used by the wider research team. Further to that, the FSM information used for the impact evaluation will be pseudonymised and therefore cannot be used to identify individual students.

Teaching-as-usual group

The other case study visits will focus on teachers randomised to the teaching-as-usual group. The unit of analysis will again be the GCSE Maths resit teacher. The aim of these case studies is to establish what is taking place instead of Mastering Maths in teaching-as usual-settings (RQ 8) and to understand wider context and moderators (RQ 11). Potential spillover effects within a setting will also be a focus. To explore spillover effects, we will ask teaching-as-usual teachers about their awareness of Mastering Maths, whether resources or ideas have been shared with them from the programme, and their level of familiarity with the Teaching for Mastery approach and Centre for Excellence in Maths research lessons.

For teaching-as-usual teachers, we will do a single one-day case study visit that includes the following activities:

- Observation of teacher's regular maths lesson – **monitoring the control**
- Student interview (no longer than 30 minutes; with students being given the option of doing these individually or in pairs) – **reach, responsiveness, monitoring the control, perceived impacts, context/moderators**

[52] In the original version of the protocol, we had planned to ask students to disclose this at the start of interviews. We changed this approach so that students were asked at the end, allowing time for the interviewer to build a positive rapport with the pupil versus asking a potentially sensitive question at the beginning of the encounter and risking their engagement with subsequent questions.

- Teacher interview (no longer than 50 minutes) – **monitoring the control, context/moderators, spillover effects**

The structured observation protocol and semi-structured topic guides for all teaching-as-usual case study activities will ask questions relating to the IPE dimension of **monitoring the control**. The student interview and teacher interview will additionally include questions relating to **context/moderators**, including questions about the college's policy towards the November resits and whether knowing that they had been allocated to the teaching-as-usual condition affected the teachers' or the students' engagement with the trial.

Endline surveys (intervention and teaching-as-usual)

We will collect survey data from all intervention and teaching-as-usual students and teachers at endline in March 2025.

Both teacher and student surveys will be paper based in an effort to boost response rates. The alternative online design would require teachers either sharing an online survey link with students (which would likely reach a very low response rate) or facilitating students completing the survey at setting (which would likely be a barrier to the teacher's willingness to facilitate the survey fieldwork).

In our paper-based design, we will courier survey packs to all teachers. Intervention and teaching-as-usual group teachers in the same setting will receive separate survey packs addressed to the individual teacher. The survey pack will include named teacher and student surveys and instructions for how to return completed surveys to NatCen. The student surveys will be pre-filled with some of the students' details, such as their full names in order to facilitate linking with the NPD and endline data during analysis. We will derive the information from the February data return from settings that will have excluded students who achieved a grade 4 in the November 2024 resits.

We will ask teachers to administer the paper surveys to students during a maths lesson to maximise participation. If any students are absent on a particular day, teachers will have several other chances to administer the survey to those students at a later date. We will ask teachers to return the completed questionnaires, including their own, to NatCen by the end of the Spring term. We will arrange a courier for collection.

The endline surveys will be different for intervention and teaching-as-usual teachers and students. For intervention teachers, the teacher survey will ask about **fidelity**, **adaptation**, **dosage**, **responsiveness**, **differentiation**, **perceived impacts**, **context/moderators** and **costs**. The intervention student survey will ask about **responsiveness**, **differentiation**, **perceived impacts**, and **context/moderators,** including the scales for the secondary outcome of student confidence and self-efficacy towards maths. For teaching-as-usual teachers, we will ask teachers questions relating to **monitoring the control** and **context/moderators**. The teaching-as-usual student survey will include questions relating to **monitoring the control**, **context/moderators** and also include the secondary outcome scales. The surveys will take 10-15 minutes to complete.

Non-respondent teachers will be chased after they return from the Easter school holidays. We will leave the survey fieldwork open until approximately the end of May, to maximise response rates.[53]

Delivery team data collection

At baseline, in the summer of 2024, we will conduct an interview with the delivery team at UoN to capture information about changes that were made to the programme following the efficacy trial, as is recommended in the EEF's IPE guidance.[54] The guidance states that it is particularly important for effectiveness trial IPEs to capture information about changes that were made to the programme following the efficacy trial, enabling it to operate at greater scale, the implications of these changes, and any variability in implementation.

At endline, in the summer of 2025, we will interview the delivery team and review any material they have collected during implementation. This will include:

- Attendance registers for teacher professional development events and lesson study groups, collected by lead teachers
- Compliance logs which capture how many Mastering Maths lessons were taught by individual teachers during the delivery period, collected by lead teachers
- Log of implementation issues

The endline interview topic guide will include questions about barriers and facilitators to delivery, as well as the IPE dimensions of **fidelity**, **dosage**, **quality**, **reach** and **responsiveness**. In the endline interview, we will also explore any causal relationships between outcomes and moderators and contextual factors.

Both interviews will take place online or by telephone.

At endline, we will also ask the delivery team to complete a cost pro-forma for the IPE dimension of **costs**.

Analysis

NatCen will record (with participant consent) all interviews, which will then be professionally transcribed by McGowan Transcriptions, one of NatCen's approved suppliers. We will code interview and observation data using the Framework approach developed by NatCen.[55] The Framework approach is a type of thematic analysis which evidences the relationship between

[53] In the original version of the protocol, we described sending unresponsive teachers the teacher survey again later in the spring to maximise response rates. This step was removed partly because the timeline for survey collection was delayed, and also to streamline data collection. This was a decision made in collaboration with the delivery team and EEF.

[54] Implementation and process evaluation guidance for EEF evaluations. EEF-IPE-Guidance-August-2022.pdf (d2tic4wvo1iusb.cloudfront.net)

[55] Ritchie et al. (2013) Qualitative Research Practice: A Guide for Social Science Students and Researchers. London: Sage Publications.

themes and anonymised cases. For this study, the coding will be deductive, meaning that it will be based on the IPE dimensions and research questions we have set out at the beginning.

The analysis will be structured around the IPE dimensions. The findings from this analysis will allow us to respond to our research questions as well as comment on the assumptions, moderators, and contextual factors for the programme logic model.

For the case studies, we will first conduct analysis within cases (i.e., between interviews and observations in each case study) to consolidate findings within the case. We will then move to comparative analysis that will consider similarities and differences between case studies of the same type (i.e., between intervention cases, between teaching-as-usual cases) and also between case types (i.e., intervention versus teaching-as-usual). We have chosen this approach because the teacher is the most important actor in the Mastering Maths intervention: their engagement with the training and lesson study groups and consequently their understanding and teaching of Mastering Maths lessons has a significant influence on the programme that the students receive. This justifies the need for **within-case analysis** (because we expect findings to be consistent within a case) and also **between-case analysis** between intervention and teaching-as-usual cases (because we expect there to be differences between those who implemented the intervention and those who did not), between intervention group cases (because we expect there to be differences between cases based on the unique approach of each teacher and the college context they operate in), and between teaching-as-usual cases (because we expect some differences in the moderating/contextual effects and also the extent to which teachers adhere to the teaching-as-usual condition).

We will analyse survey data using SPSS. This data will be analysed using descriptive statistics.

We will analyse the results of the student surveys on the ONS' SRS. This is because we will use students' ever-FSM status obtained from the NPD in our analyses. This will help answer research question 9a on whether **perceived outcomes** differ between disadvantaged and more advantaged students. (It is worth noting here that due to a lack of a better alternative, we do not use this definition of disadvantage in the qualitative interviews with students. Instead, we ask students to self-disclose at the end of the interview whether they have ever been in receipt of FSM. This means that the qualitative findings from students will not be directly comparable with the findings of the student surveys.)

Data analysis will take place sequentially. We will first analyse qualitative interviews and observations in the summer of 2025, and then the survey data in autumn 2025 (this is the earliest that we will have access to the NPD data).

Results from analyses of surveys, interviews and observations will be triangulated (i.e., the interview and observation Frameworks will be compared with one another as well as with the descriptive and cross-tabulated survey results, in order to see how findings from different data sources are complementary or contrasting). The findings will be assessed against the IPE dimensions and research questions as well as the logic model to ascertain the extent to which short- and long-term outcomes have been achieved, and which contextual or moderating factors influence intervention delivery and impacts.

Quality assurance

To ensure data collection is carried out to the highest standards, all researchers on the project will be experienced in data collection in educational settings. All data collection materials (topic guides, observation protocols and surveys) will be quality-assured by senior members of the team. Before interviews and observations take place, researchers will be briefed on each topic guide/observation protocol separately and in detail. Researchers will be instructed not to deviate from the instruments in order to ensure consistency across the qualitative research encounters. The topic guides, observation protocols and surveys will be added as appendices upon submission of the final report, for transparency and in order to ensure replicability of findings. In addition, the method of analysis will be thoroughly described in the final report, but we anticipate that it will be an iterative process, with researchers utilising NatCen's Framework approach as described above. All analysis steps will be quality-assured by senior members of the team, with differences in findings being approached and resolved during team analysis meetings. This multi-step method of quality assurance also hopes to reduce the risk of researcher bias, due to the differing perspectives of each member of the team involved.

Table 7 provides an overview of the number of participants and encounters for each IPE activity.

Table 7: IPE methods overview

Research methods	Data collection methods	Number of encounters and participants	Data analysis methods	Research questions addressed	IPE dimension
Training observations of lead teachers	In person; structured observation	Encounters: 2 observations Participants per encounter: 15	Qualitative; framework analysis	1, 2, 4, 6	Fidelity, adaptation, quality, responsiveness
Training observations of intervention teachers	In person; structured observation	Encounters: 2 observations Participants per encounter: 5-6	Qualitative; framework analysis	1, 2, 4, 6	Fidelity, adaptation, quality, responsiveness
Lead teacher interviews	Online; semi structured topic guides	Encounters: 2 interviews (post-training and at endline) x 6 Participants per encounter: 1	Qualitative; framework analysis	1, 2, 3, 5, 6, 9, 11	Fidelity, adaptation, dosage, reach, responsiveness, perceived impacts, context/moderators
Case studies (intervention)	In person; structured observations and semi-structured topic guides	Encounters: 1 observation of teacher's lesson study group, 1 observation of teacher's maths lesson, 1 teacher interview, and 1 paired student interview Participants per encounter: 3	Qualitative; framework analysis	1, 2, 4, 5, 6, 7, 9, 11	Fidelity, adaptation, quality, reach, responsiveness, differentiation, perceived impacts, context/moderators
Case studies (teaching-as-usual)	In person; structured observation and semi-structured topic guides	Encounters: 1 observation of teacher's maths lesson, 1 paired student interview and 1 teacher interview Participants per encounter: 3	Qualitative; framework analysis	8, 11	Reach, responsiveness, monitoring control, context/moderators
Endline surveys (intervention)	Teacher survey (in person/online) Student survey (in person)	Encounters: 2 endline surveys Participants per encounter: 1,000+	Quantitative; descriptive statistics	1, 2, 3, 5, 6, 7, 9, 10, 11	Fidelity, adaptation, dosage, reach, responsiveness, differentiation, perceived impacts, costs, context/moderators
Endline surveys (teaching-as-usual)	Teacher survey (in person/online)	Encounters: 2 endline surveys Participants per encounter: 1,000+	Quantitative; descriptive statistics	8, 11	Monitoring control, context/moderators

	Student survey (in person)				
Delivery team data collection	Online or telephone; semi structured topic guide; cost proforma	Encounters: 2 paired interviews, 1 review of materials (attendance registers for teacher professional development events and lesson study groups, compliance logs, log of implementation issues), 1 cost proforma Participants per encounter: 2	Mixed methods; framework analysis and descriptive statistics	1, 3, 4, 6, 10	Fidelity, dosage, quality, reach, responsiveness, costs

Cost evaluation design

We will collect costs data consistent with EEF guidelines[56]. We will collect data from:

1) Endline teacher surveys to calculate the cost of *attending* professional development and lesson study groups as well as *preparing* for and *teaching* Mastering Maths lessons and usual practice maths lessons
2) A proforma completed by the delivery team on the costs incurred to the delivery team, including personnel costs for preparing programme delivery personnel costs during training and implementation of the programme, as well as facilities, equipment and materials for implementation.[57]

Following EEF guidance, we will estimate the total cost per setting for the intervention as implemented for three consecutive years, categorising costs into pre-requisites as well as start-up and recurring costs. The cost per-pupil-per-setting-year will be based on the total cost per setting divided by the number of students per setting-year expected to benefit from the intervention. We will use the actual number of students from the trial.

Given possible heterogeneity in costs and parameter uncertainty across settings, we will also report costs for different ingredients as illustrated in Table 8 (including statistics such as the mean, median, minimum, maximum and standard deviation). We will consider using extreme values for ingredients to create best/worst cases in the event of more outliers than the normal distribution.

Table 8: Potential resources and evaluation sources for cost evaluation

Category (ingredient)	Item	Evaluation data sources	Scope
Personnel for training	Lead teachers prepare and deliver professional development and lesson study groups	Delivery team proforma	All settings within intervention group
	Teachers attend professional development and lesson study	Teacher survey (to include a question about whether/how	

[56] Cost evaluation guidance for EEF evaluations. /https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/Cost-Evaluation-Guidance-Feb_2023.pdf?v=1711356652

[57] The proforma was a later addition. In the original protocol, we had planned to collect cost data in lead teacher interviews and the delivery team endline interview. We changed the approach away from lead teachers as they were not best placed to answer cost questions, For the delivery team, a proforma was preferred over the interview as it allowed them more time to collate the necessary information.

	groups, within teachers' directed time – no extra cost anticipated	teachers are paid or reimbursed for training and meetings)	
Personnel for implementation	Teachers prepare and teach Mastering Maths lessons, within teachers' directed time – no extra cost anticipated	Teacher survey	All settings within the intervention group
Training and programme-level costs	Delivery team develop training materials for lead teachers	Delivery team proforma	Delivery team
	Lead teachers attend 'Train the trainer' training developed by the delivery team	Delivery team proforma	
Facilities, equipment and materials	Additional resources needed for Mastering Maths lessons – no extra cost anticipated	Teacher survey, Estimation using market price	Estimation for all settings with intervention group
Business-as-usual costs	Usual costs of running maths lessons	Teacher survey	All settings
	Actual costs of maths lessons in teaching-as-usual condition during trial period	Teacher survey	All settings within the teaching-as-usual group

Ethics and registration

NatCen has a robust ethics procedure. The Research Ethics Committee (REC) is designed to provide ethical guidance and advice, and to ensure all research undertaken by NatCen will meet the ethical standards of government and other funders. NatCen's ethics procedure is aligned with the 2021 Government Social Research (GSR) professional guidance and the 2021 Social

Research Association (SRA) ethics guidance. For this project, an ethics application was submitted to the REC on 5th January 2024, receiving ethical approval on 11th January 2024.

Agreement to participate in a trial

Recruitment of settings

UoN will be responsible for recruiting settings into the trial. They will aim to recruit approximately 70 college settings across England starting in February 2024. In collaboration with UoN and the EEF, NatCen have drafted the trial recruitment documents, including a setting information sheet, a memorandum of understanding (MoU), a student information sheet, and a privacy notice.

The setting information sheet includes an overview of what Mastering Maths is, the concept of random assignment, the aims of the independent evaluation, potential benefits and costs for settings taking part, a timetable of planned activity, and information regarding data protection and ethical approval. The MoU explains in more detail the expectations and responsibilities for teachers in both the intervention and teaching-as-usual groups. In all settings, the participating teacher(s) and a senior leader will be asked to sign the MoU on behalf of the setting. By doing so, settings will be formally agreeing to take part in the evaluation and to the associated conditions.

Recruitment of teachers

Eligible teachers will be those who are teaching a post-16 GCSE Maths resit class, but who have not taken part in the full Mastering Maths efficacy trial or plan to take part in a programme that tries to achieve the same outcomes in the same sector. Two teachers from each of the approximately 70 college settings will be recruited and (as detailed above) will sign the MoU to indicate their agreement with taking part. The aim is for 140 teachers to be recruited, with a maximum of two teachers being eligible from a single college setting. Settings with only one teacher are still able to participate.

For the IPE activities, NatCen will be responsible for recruiting teachers. We will receive a list of settings from UoN that includes details of intervention and teaching-as-usual teachers. We will use this list to send postal surveys to teachers. We will liaise with lead teachers to select teachers for case studies. The teachers who are selected will be invited to take part but are under no obligation to do so. They will receive a further information sheet explaining the aim of the research. We will obtain verbal consent from participants at the start of interviews.

Recruitment of students

Eligible students will be those aged 16-19 who are enrolled in the classes of participating teachers and are resitting their GCSE Maths exam as a condition of funding for their education. All eligible students will be given an information sheet which explains the aim of the research, how their information will be used, and their right to opt out of the study before their personal information is shared with NatCen. They will also be given a privacy notice explaining how their data will be used. Students will be able to withdraw from the evaluation at any point. It will be made clear that opting out will not affect the maths teaching that they receive.

A further information sheet will be shared with students of teachers who are selected as case studies closer to the time; we will also obtain verbal consent from students at the start of the interview. Because students are aged 16 or over, parental consent is not required. However, participants will be subject to the same NatCen safeguarding and disclosure policies and thresholds as any project involving respondents under the age of 18.

Trial registration

The trial was registered on the Open Science Framework (OSF) in September 2024. The registration link is: https://osf.io/chwkf/.

Data protection

For the duration of this evaluation, NatCen will be the data controller. Our legal basis for processing the data is 'legitimate interest'. This means that we believe there is a genuine reason for us to process this data (to evaluate the effectiveness of the Mastering Maths programme), that this data is needed to fulfil this purpose (we could not evaluate the Mastering Maths programme without this information), and that using this data will not interfere with individuals' interests, rights or freedoms.

NatCen will collect primary data from the delivery team, setting staff and students for the purposes of the IPE and the IE. NatCen will also access data held on the NPD using student identifiers (Unique Learner Numbers) received from teachers.

NatCen will receive personal data relating to students and setting staff directly from settings as well as from the delivery team at UoN. For any data the UoN collect, they will be the data controller. There is more information about UoN's approach to data protection in their privacy notice: nottingham.ac.uk/utilities/privacy/privacy.aspx

All personal data will be transferred using a secure NatCen server. This includes the students' raw GCSE scores in the summer and autumn of 2025.

The data subjects will be students, setting staff and the delivery team.

NatCen will store and handle all data securely and confidentially, in line with UK GDPR and the Data Protection Act 2018. Only the research team at NatCen will have access to data collected as part of the evaluation. This will be monitored through a data security plan set up by the Principal Investigator. In addition, some third parties (e.g., transcription company, printing company for surveys) will have limited access to personal data. There is more information about who has access to data collected as part of this evaluation on NatCen's privacy notice: natcen.ac.uk/MasteringMaths-Privacy

Any reports or publications arising from this research will not identify any individual research participant or setting. All personal information, and any other data held on the project, will be securely deleted from NatCen systems within twelve months of the project's completion. The end date of the project is estimated to be May 2026, meaning deletion will take place in May 2027.

Data collected as part of all EEF evaluations are archived. During the evaluation, NatCen will share the impact evaluation data with the DfE in order for it to be linked to the NPD. After the evaluation, NatCen will share the data with the EEF for archiving. Following this, EEF will be the data controller of the archived data. For further information, see the EEF's archive privacy notice.

Personnel

A team of research and evaluation specialists at NatCen will carry out the evaluation. The Centre for Children and Families will lead the evaluation with support from impact evaluation specialists from the Centre for Evaluation.

Table 9: Evaluation team

Evaluation team	Institution	Role
Helena Takala	Research Director, Centre for Children & Families	Principal Investigator
Dr Enes Duysak	Research Director, Centre for Evaluation	IE Lead
Abigail Rennick	Senior Researcher, Centre for Children & Families	Day-to-day project manager and IPE Lead
Tien-Li Kuo	Senior Researcher, Centre for Evaluation	IE support
Charlotte Tomlinson	Researcher, Centre for Children & Families	IPE support
Charlotte Bessant	Researcher, Centre for Children & Families	IPE support
Dr Gayle Munro	Director, Centre for Children & Families	IPE quality assurance
Dr Terry Ng-Knight	Research Director, Centre for Evaluation	IE quality assurance

A team from UoN will be responsible for the delivery of the Mastering Maths programme.

Table 10: Delivery team

Delivery team	Institution	Role
Professor Geoffrey Wake	Professor of Mathematics Education, University of Nottingham	Director
Dr Marie Joubert	Senior Research Fellow, University of Nottingham	Professional development lead and key liaison with lead teachers
Dr Michael Adkins	Senior Research Fellow, University of Nottingham	Statistical analysis support
Corrine Robinson	Senior Operations Officer, University of Nottingham	Day-to-day administration and support for team

Risks

Table 11: Risk management

Risk	Likelihood/ Impact	Mitigation/Contingency
Not receiving student lists during November 2024 from settings	**Likelihood:** Medium **Impact:** High	• During recruitment, settings will put forward one to two named teachers for the study, who will sign an MoU that states their expected roles and responsibilities, including providing student lists in autumn 2024. • We will communicate with teachers about this task early on, to allow them time to prepare resource to gather data and set up their accounts, which will allow for lists to be submitted on time. • We have not included baseline surveys in the design in an effort to reduce burden on college staff in autumn 2024 when their help is needed with student lists. • We have allowed for significant resource to offer any assistance to teachers, and to chase them before the deadline. • Our secure site for submitting student lists is tried and tested from previous EEF trials and designed to be as user-friendly and accessible as possible. • Settings that nominate two teachers for the trial will receive a financial incentive of £500 as a thank you for the participation of the teaching-as-usual teacher and for the collection of student data. Settings that nominate one teacher, who is then randomised to the teaching-as-usual group, will receive a financial incentive of £1,000. These payments are budgeted for and administered by the delivery team and will be paid once all data has been supplied. • We will work closely with the delivery team at the UoN to follow up with any unresponsive settings.
Not receiving the updated student lists in early 2025 from settings on time	**Likelihood**: Low **Impact**: High	• Teachers will already have their login details to the secure server from the November 2024 uploads, so there is no need to set up this process again. • Updating the student lists in early 2025 will be a simpler task than the original submission in November 2025, as teachers will just need to update the existing spreadsheet with additional details. • NatCen will have established contact with the one to two named points of contact, aiding with communication with settings.

		• We will work closely with the delivery team at UoN to follow up with any unresponsive settings. • We will communicate with settings about this task early on, to allow settings time to prepare sufficient time and resource to gather data and set up their accounts, which will allow for lists to be submitted on time. • We have reserved significant researcher time for this task, should this be needed.
Not receiving raw GCSE scores from settings on time (aiming for this to take place in August 2025)	**Likelihood:** Medium **Impact:** High	• NatCen will collect the contact details of all data managers (with UoN's support) and will have contacted them in advance of data submission to ask them to set up their log in details for the secure site. This should enable timely transfer of data. • We will work closely with the delivery team at UoN who have extensive experience with this type of data collection from the efficacy trial. • We have also allowed significant researcher time for this crucial step, to help data managers with any queries and to chase them to submit their data on time. • As mentioned above, there will be a financial incentive of £1,000 to settings that nominate one teacher who is randomised to the teaching-as-usual group, and £500 to settings that nominate two teachers, as a thank you for the participation of the teaching-as-usual teacher and for supplying data. These will be paid once all data has been supplied. • Should we not receive the scores, we have accounted for high attrition rates in our power calculations.
Mismatch between NPD data and setting data	**Likelihood:** Low **Impact:** High	• We will first double check with the setting whether the information they share with us is correct. • If the mismatch is confirmed, the observation will be removed from the sample for the outcomes of interest. Note that the attrition accounted for in the power calculations includes such mismatches in similar levels as those observed during the efficacy trial.
Limited teacher and college engagement with	**Likelihood:** Medium **Impact:** High	• We will work with the delivery team to ensure their own quality-assurance surveys and visits do not clash with ours.

case studies and surveys		• We have not included baseline surveys to avoid overburdening staff with research activities. • We have only included eight case studies in the design, which we believe is sufficient to answer our IPE research questions, while keeping the burden on staff at a minimum. • The timing of case study visits (February 2025) has been carefully thought of to ensure it does not clash with the teacher professional development and submission of student lists (November 2024 and February-June 2025) or the endline surveys (March to April 2025) and submission of raw GCSE scores (August to October 2025). • For those teachers participating in case studies, NatCen will administer financial incentives to the value of £200, as a thank-you for their participation and engagement. • We have chosen to administer paper-based surveys with teachers and students, which based on previous experience we expect to gain a higher response rate in comparison to an online only approach. Non-responsive teachers will be chased with a second round of paper survey fieldwork in the spring/summer.
Setting attrition	**Likelihood:** Medium **Impact:** High	• MoUs will outline details of roles and responsibilities for participating settings so that they can make an informed decision about joining the study. • We have designed the evaluation to minimise burden on settings while ensuring depth of data; the key evaluation activities are spaced across the academic year, to minimise burden on college staff and to ensure their consistent engagement. • A financial incentive of £1,000 for settings who nominate one teacher who is then randomised to the teaching-as-usual group, paid towards the end of the evaluation, will encourage engagement for the duration. This could serve as a contribution towards the associated costs of the programme if they would like a teacher to participate in 2025/26. They will also be granted priority access to the programme during the subsequent academic year (2025/26). • A financial incentive of £500 for settings who nominate two teachers, plus the delivery of significant teacher professional development, will encourage engagement for the duration of the trial.

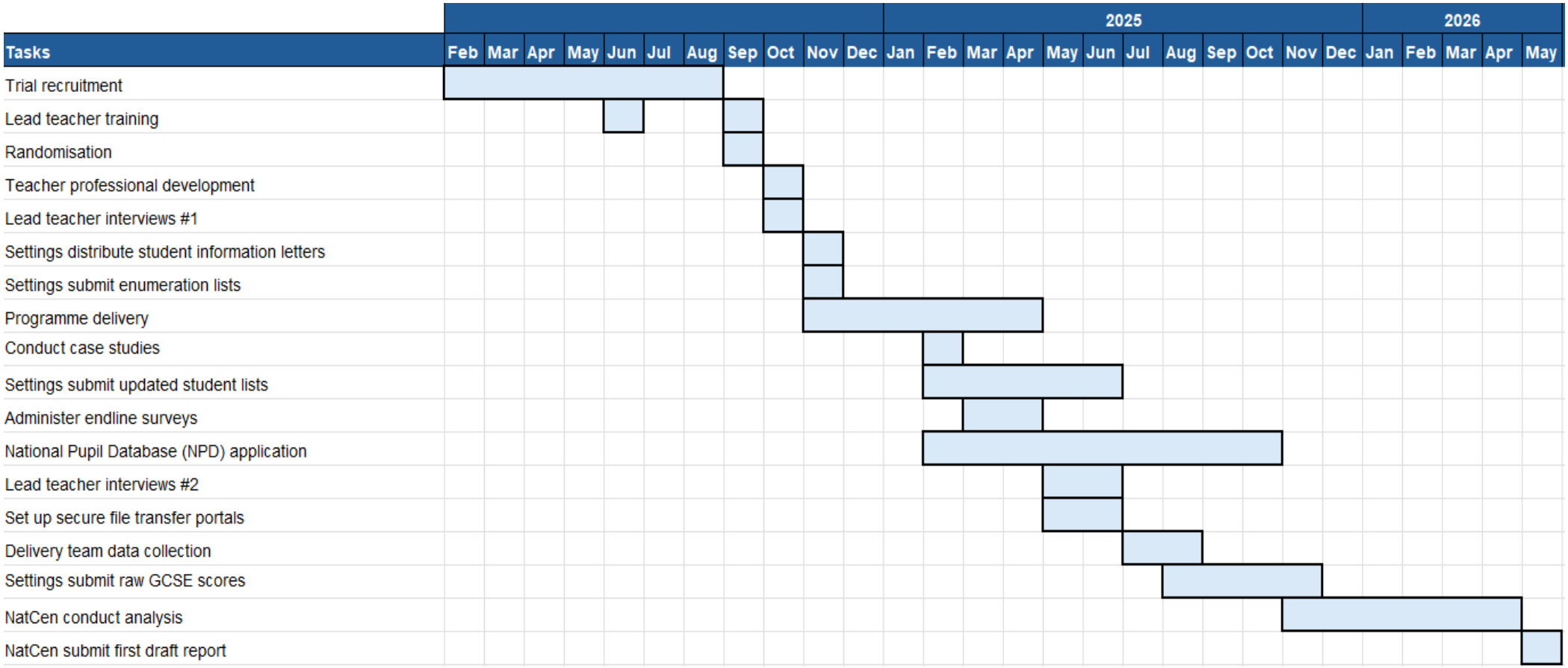
Student and staff level attrition	**Likelihood:** Medium **Impact:** High	• We have accounted for high levels of attrition in our power calculations.
Evaluation team member absence or turn-over	**Likelihood:** Medium **Impact:** Low	• NatCen staff have a three-month notice period to allow time for handover. The team can be supplemented by researchers with relevant topic and methodological expertise from a pool of researchers.

Timeline

Table 12: Timeline

Dates	Activity	Staff responsible/ leading
Feb – August 2024	Trial recruitment	UoN
June 2024	First half of lead teacher training	UoN
September 2024	Randomisation	NatCen
August – Sept 2024	Delivery team interview	NatCen
Sept 2024	Second half of lead teacher training	UoN
Oct 2024	Two days of teacher professional development	UoN
Oct 2024	Lead teacher interviews #1	NatCen
Nov 2024	Student letters distributed by settings and initial enumeration of eligible students	NatCen / Settings
Nov 2024 – Apr 2025	Programme delivery (running of lesson study groups by lead teachers and of Mastering Maths lessons by teachers)	UoN / Settings
Feb 2025	Case studies	NatCen
Feb - June 2025	Updated student lists	Settings
Mar – Apr 2025	Endline surveys, including student confidence and self-efficacy measures	NatCen / Settings
Feb – Oct 2025	NPD application	NatCen
May – June 2025	Lead teacher interviews #2	NatCen
May – June 2025	Set-up of secure file transfer portals (FTPs) with settings	NatCen
Jul – Aug 2025	Delivery team data collection (including compliance data)	NatCen
Aug – Nov 2025	Collection of raw GCSE scores	NatCen / Settings
Nov 2025 – April 2026	Analysis	NatCen
May 2026	First draft report submitted	NatCen

Figure 3: Visual timeline



References

Association of Colleges. (2020). *College Key Facts 2019/20*. https://d4hfzltwt4wv7.cloudfront.net/uploads/files/AoC-College-Key-Facts-2019-20.pdf

Camden, B. (2023). *GCSE resits 2023: Maths and English pass rates down again*. FE Week. https://feweek.co.uk/gcse-resits-2023-maths-and-english-pass-rates-down-again/

Crips, B., Hallgarten, J., Joshua, V., Morris, R., Perry, T., & Wardle, L. (2023). *Post-16 GCSE Resit Practice Review*. Centre for Education & Youth, and University of Warwick. https://d2tic4wvo1iusb.cloudfront.net/production/documents/Post-16-GCSE-Resit-Practice-Review.pdf?v=1689345688

Dysart, E., Hanley, P., Coleman, E., Tracy, L. (2023) *5Rs Retrial Evaluation Protocol*. Education Endowment Foundation. https://d2tic4wvo1iusb.cloudfront.net/documents/projects/V1_5Rs-EEF_trial_protocol_2022.docx_CLEANMarch2023.pdf?v=1682079782

Education Endowment Foundation [EEF]. (2022) *Statistical analysis guidance for EEF evaluations*. *https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1696581237*

Education & Training Foundation. (n.d). *CFEM Mastery Principles*. https://www.et-foundation.co.uk/professional-development/maths-and-english/cfem/cfem-resources-and-evidence/teaching-for-mastery-classroom-resources/cfem-mastery-principles/

GOV.UK. (2014). *Independent report: Review of vocational education: The Wolf report*. Department for Education, and Department for Business, Innovation and Skills. https://www.gov.uk/government/publications/review-of-vocational-education-the-wolf-report

GOV.UK. (2024). *16 to 19 funding: maths and English condition of funding*. https://www.gov.uk/guidance/16-to-19-funding-maths-and-english-condition-of-funding#:~:text=To%20meet%20the%20condition%20of%20funding%2C%20a%20student%20must%20study,as%20'stepping%20stone'%20qualifications.

Hanley, P., Elliott, L., Fairhurst, C., & Coleman, E. (2021). *The 5Rs approach to GCSE Maths resits: Trial Evaluation Protocol*. https://d2tic4wvo1iusb.cloudfront.net/production/documents/pages/projects/5Rs_post-16_GCSE_Resit_maths_COVID_clean.pdf?v=1696581772

House of Commons Education Committee. (2023). *Oral evidence: Government Proposals for Compulsory Maths to Age 18*. https://committees.parliament.uk/oralevidence/12668/pdf/

Humphrey, N., Lendrum, A., Ashworth, E., Freason, K., Buck, R. and Kerr, K. (2019) *Implementation and process evaluation (IPE) for interventions in education settings: an introductory handbook*. London: EEF.

Imbens, G., Angrist, J. (1994) *Identification and Estimation of Local Average Treatment Effects*. https://business.baylor.edu/scott_cunningham/teaching/imbens--angrist---late-1994.pdf

Maughan, S., Smith. J., Mitchell, T., Horrocks, N., & Taylor, A. (2016). *Improving Level 2 English and maths outcomes for 16 to 18 year olds – Literature review*. Education Endowment Foundation. https://d2tic4wvo1iusb.cloudfront.net/production/documents/guidance/16-18_Literature_Review.pdf?v=1695718463

Noyes, A., & Dalby, D. (2020). *Mathematics in England's Further Education Colleges: an analysis of policy enactment and practice*. Nuffield Foundation. https://www.nottingham.ac.uk/research/groups/crme/documents/mifec/interim-report-2.pdf

Kuczera, M., Field, Simon., Windisch, HC. (2016) *Building Skills for All: A Review of England*. OECD. https://www.oecd.org/unitedkingdom/building-skills-for-all-review-of-england.pdf

Ritchie, J., Lewis, J., Nicholls, C.M. and Ormston, R. (2013). *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. 2nd ed. London: SAGE Publications Ltd.

Smith, A. (2017). *Smith review of post-16 mathematics: Report and letter*. GOV.UK. https://www.gov.uk/government/publications/smith-review-of-post-16-maths-report-andgovernment-response

The Royal Society. (2022). *GCSE Resits Project: background*. https://royalsociety.org/-/media/policy/topics/education-skills/gcseresits/GCSE-resits-project-background.pdf

Wake, G., Adkins, M., Dalby, D., Hall, J., Joubert, M., Lee, G., & Noyes, A. (2023). *Centres for Excellence in Maths Teaching for Mastery Randomised Controlled Trial Evaluation Report*. https://www.nottingham.ac.uk/research/groups/crme/documents/cfem-tfm-report.pdf

Education & Training Foundation. (n.d). *Centres for Excellence in Maths Mastery Principles*. https://www.et-foundation.co.uk/professional-development/maths-and-english/cfem/cfem-resources-and-evidence/teaching-for-mastery-classroom-resources/cfem-mastery-principles

Appendix A: Changes since the previous evaluation

Appendix Table 1: Changes since the previous evaluation

Feature		Efficacy to effectiveness stage
Intervention	Intervention content	The efficacy trial included a partial intervention, where teachers received two professional development days (two half-days and one full day) and five sets of teaching materials, and a full intervention, where teachers received three full professional development days in addition to five local cluster meetings led by a lead teacher. This trial only includes the full intervention.
	Delivery model	The delivery mechanism has changed from a developer-led training to a train-the-trainer type model. In the efficacy trial, the developers ran the initial professional development and lead teachers ran the cluster meetings, whereas in the current trial, lead teachers run all of the professional development. In addition, during the efficacy trial, the initial training and some cluster meetings were held online due to Covid. For this trial, delivery is planned to be entirely face-to-face.
Evaluation	Eligibility criteria	There are no changes in students' eligibility criteria. However, teacher and setting eligibility criteria has changed. Teachers are not eligible to take part in the new trial if they took part in the full intervention arm of the efficacy trial, in the NCETM's Teaching for Mastery programme, or are taking part in programmes involving substantial professional development during the 2024/25 academic year.
	Level of randomisation	The level of randomisation has changed from setting level in the efficacy trial to teacher level in the effectiveness trial. This means that there may be both an intervention and control teacher in the same setting for this trial, which was not the case for the efficacy trial.
	Multi-arm RCT design	In the efficacy trial, a three-arm cluster randomised controlled trial was conducted (full intervention, partial intervention, and teaching-as-usual group), while in the effectiveness trial, we are conducting a two-arm cluster randomised controlled trial where teachers are randomised into either the Mastering Maths intervention or the teaching-as-usual group.
	Outcomes and baseline	While the efficacy trial had the secondary outcome of Mastery subscale and the GCSE Maths grade score, the effectiveness trial focuses on students' probability of achieving grade 4 or higher and their probability of moving up a grade. In the effectiveness trial, we also have an additional secondary outcome to measure student attitudes towards maths, using self-confidence

		and self-efficacy measures scales. There are no changes to the baseline measures and the primary outcome.
	Control condition	No changes