CENTRES FOR EXCELLENCE IN MATHS

University of Nottingham
UK | CHINA | MALAYSIA

Report

Centres for Excellence in Maths Teaching for Mastery Randomised Controlled Trial

Evaluation Report

Geoffrey Wake
Michael Adkins
Diane Dalby
Jonathan Hall
Marie Joubert
Gabriel Lee
Andy Noyes

March 2023

Table of contents

About the evaluators

The project was independently evaluated by a team from the University of Nottingham Education: Geoff Wake, Michael Adkins, Diane Dalby, Jonathan Hall, Marie Joubert, Gabriel Lee, Andy Noyes.

The lead evaluator was Geoffrey Wake.

Contact details:
Professor Geoffrey Wake
School of Education
University of Nottingham
Wollaton Road
Nottingham NG8 1BB
Tel: +44 (0) 115 846 6219
Email: geoffrey.wake@nottingham.ac.uk

Note:

Acknowledgements

The project team would like to acknowledge the cooperation and hard work of the many teachers without whom the Teaching for Mastery in FE randomised controlled trails would not have been possible. The Lead Teachers and Trial Teachers' enthusiasm for, and commitment to, the programme was inspiring. Their response provides evidence, if it were needed, of the dedication of teachers in the sector to helping students who need much care and support to overcome their previous underachievement and disappointment in mathematics. We especially extend our thanks to the students who were part of this research. Their engagement in the classroom and their desire to overcome previous lack of understanding was very much visible and their efforts to do so was truly uplifting.

We are particularly grateful to colleagues in case study colleges who gave access to their classrooms and the students themselves, without whom the detailed insights that the research affords would not have been possible.

The team would also like to thank colleagues at the Education and Training Foundation (ETF), who were lead partners of the CfEM programme, the Department for Education (DfE) as funders, the Centre Leads and other colleagues of the Centres for Excellence in Maths programme for their assistance in working with their network partners to recruit to the trials. The national trials were a massive undertaking and a challenge to design in nine months. We are grateful to all involved for the parts they played and the assistance they gave to the research team at all stages of development and implementation. Equally, we wish to acknowledge the teachers and students who went about their business as usual and who provided the data that allowed us to determine the impact of the intervention. Without their participation none of this would have been possible.

Finally, as a research team we work closely with our Research Operations Team in the School of Education and without their careful attention to the detail of collecting and processing the large volume of data from across all colleges we could not have managed to complete this work. We would particularly like to acknowledge the work of Kanchana Minson, Alex Phillips, and James Fox who always managed the work of the project team with dedication and good humour.

Executive summary

The project

The intervention was designed to improve post-16 students' learning of level 2 mathematics. In general, the approach used was one of teaching for mastery in FE, aligned to principles developed by teachers and educators working with the Centres for Excellence in Maths programme. A Handbook sets out these principles alongside advice relating to teaching in this way.

The intervention targeted teachers of students undertaking post-16 GCSE mathematics resit classes. The aim was to support the teachers to develop new teaching practices aligned with the Mastery Teaching principles by engaging with professional development that explained the approach and working with seven lessons that exemplified this. Five of the lessons were taught in specific time frames over the duration of the course using these to inform their developing teaching approach. There were two levels of intervention investigated: a partial intervention that included all the aforementioned and a full intervention that in addition included a programme of a modified version of lesson study. The lesson study was led by a small cohort of Lead Teachers who had taken part in an earlier pilot and who had some additional online training. Teachers in this arm of the trial were clustered geographically to take part in the lesson study process.

The evaluation was a three-armed randomised controlled efficacy trial, involving 147 colleges and 7453 students. The process evaluation included teacher surveys and 13 case studies. Recruitment to the trial was managed by Centre Leads of the CfEM and drew on teachers from Centres and their wider partner networks. The intervention ran in colleges from October 2021 to June 2022. During this period illness due to Covid-19 and college-imposed visiting restrictions proved an issue for teachers and students resulting in some PD sessions and cluster meetings being held online instead of face-to-face as originally planned. Covid-19 also led to some attrition of teachers from the trial as workload proved difficult due to illness of participating teachers and their colleagues.

Table 1 Key conclusions

Key Conclusions
1. GCSE resit students taught by teachers in the full Mastery Teaching intervention made one month additional progress in mathematics learning compared to students in other (business as usual) colleges.
2. Students having had Free School Meals prior to college, and taught by teachers participating in the full Mastery Teaching intervention, made two months additional progress in mathematics learning compared to students in other (business as usual) colleges.
3. Teachers in both intervention groups report that taking part in the PD intervention programme and teaching the exemplary Teaching for Mastery lessons as: ■ being effective as an introduction to the principles of Teaching for Mastery ■ leading to their improved understanding of how to implement Teaching for Mastery in their practice ■ leading to changes in their teaching practice during the programme and high levels of intended change in teaching practice (in subsequent years) ■ resulting in improved student engagement and understanding.
4. Compliance was generally fair, with close to two-thirds of settings in both arms of the intervention teaching lessons 1-4. Compliance decreased for both arms in the teaching of lesson 5 with over half of teachers in the full intervention, but only just over two-fifths of the partial intervention teachers teaching the final lesson. There was high fidelity in terms of the lesson aims and design in the teaching of the sample lessons and teachers reported trying to implement the Teaching for Mastery approaches in their other lessons.

Additional findings

At a more detailed level we investigated a secondary outcome impact measure based on a sub-scale score for GCSE questions that aligned with the content of the exemplary Teaching for Mastery lessons taught by teachers in both full and partial intervention groups.

Analysis confirms that of the primary outcome measure and detects a slightly greater impact on (FSM) students taught by teachers in the full intervention. This again gave an effect size (of 0.13) that suggested two months of additional progress.

The intervention was delivered as intended. Lead teachers report being highly engaged with the training sessions that prepared them for working with the Trial Teachers and increasing their knowledge over the course of the phased PD of the mastery approach, the exemplary lessons and the lesson study approach. Likewise, the Trial Teachers report their growing understanding of the mastery approach over the course of the year. Trial Teachers worked hard to follow the mastery approach within the trial lessons, although in some cases this entailed making significant changes to their teaching, such as allowing students to struggle towards understanding. Most made only minor changes to the lesson plans to take account of individual

classes they were teaching. The teachers reported that they were increasingly using these approaches in their other lessons, mainly in terms of spending more time on whole class discussion and extended pair work.

The intervention took place during the course of the academic year 2021-22 with the GCSE examination taking place in June 2022 for the first time in three years because of disruption due to the Covid-19 pandemic. The pandemic led to two issues that should be considered in relation to the implementation and results of the intervention.

First, the students 'resitting GCSE' had not in fact taken a GCSE examination previously and had experienced a very disruptive period of learning in Years 10 and 11 prior to being at college. This impacted on the cohort of students in the study in a number of, predominantly negative, different ways. In particular, their motivation and engagement with learning was potentially negatively impacted and the cohort as a whole was likely to have been less well-prepared mathematically than previous cohorts.

Second, illness due to covid was particularly disruptive over the winter months of the intervention period. The original design of the intervention involved face-to-face PD meetings for Lead Teachers and Trial Teachers. Lesson Study style cluster meetings for teachers in the full intervention were also planned to be face-to-face (an important part of the process). Unfortunately, all PD sessions were held online because of covid restrictions, often imposed by colleges, and due to illness of teachers and colleagues. Likewise, two of the planned five cluster meetings were also held online. Contingency plans had been put in place in preparation of the intervention in the likely event that this was going to happen. These plans were put into operation.

Cost

The average cost of Maths Mastery in FE for one setting was around £4,107.5 or £50.02 per student per year when averaged over three years. This assumes 30 students per year, rising cumulatively from 30 students in the first year to 90 students in the third year, however, given the condition of funding requirement, there will be students resitting GCSE Mathematics several times. We have not adjusted the costs to take account of this. We would expect the costs to be further reduced in a broader scale up. In the materials only arm, the costs per student averaged over three years was estimated at £18.92.

Impact

Table 2 Summary of impact on primary outcome, and primary and secondary contrasts (Source: ONS)

Outcome & Group	Effect Size (95% Credible Interval)	Estimated months' progress[1]	No of pupils observed (intervention; control)
GCSE Mathematics standardised raw score (z-score by board) Full intervention vs. Control[2]	0.06 (-0.12, 0.24)	1	2516 (889, 1627)
GCSE Mathematics standardised raw score – Free School Meal (FSM) students Full intervention vs. Control[2]	0.11 (-0.10, 0.32)	2	903 (323, 580)
GCSE Mathematics standardised raw score (z-score by board) Partial intervention vs. Control[3]	0.04 (-0.16, 0.25)	0	2501 (874, 1627)
GCSE Mathematics standardised raw score – Free School Meal (FSM) students Partial intervention vs. Control[3]	0.03 (-0.19, 0.25)	0	896 (316, 580)

[1] Estimated months' progress is based on effect sizes reported in British and international research studies. For more information, see https://educationendowmentfoundation.org.uk/education-evidence/using-the-toolkits; [2] Primary contrast, see the results section for more information; [3] Secondary contrast, see the results section for more information

Introduction

The Centres for Excellence in Mathematics (CfEM) programme is a national improvement project aimed at delivering a step change in mathematics teaching up to Level 2 in post-16 settings. It involves 21 Centres for Excellence in Mathematics (predominantly General Further Education Colleges (GFECs)) and a consortium of expert delivery partners, managed and led by the Education and Training Foundation (ETF).

The programme activities in 2019/2020 included a set of pilot research trials, the initial stages of the development of local college networks and action research projects situated in the Centres. In this period the University of Nottingham (UoN) Centre for Research in Mathematics Education (CRME) conducted four pilot research trials (October 2019 – April 2020). In each trial, teachers used a different approach to teaching mathematics with students studying to resit their GCSE examinations. The four themes explored teaching for Mastery, using contextualisation, working with technology, and improving motivation and engagement). In March 2020 national measures were taken to combat the Covid-19 pandemic. This closed schools and colleges and eventually led to the cancellation of GCSE examinations. This meant that all work on the pilot research trails ceased, and the research team was only able to report on a much truncated set of outcomes of these pilots.

Covid-19 continued to impact substantially on education nationally, including colleges, throughout much of the following year and GCSE examinations were again cancelled in 2021. During that period the CRME research team continued to work building on the experiences of the pilot trials to develop an intervention based on Teaching for Mastery (TfM) and designed and prepared for a randomised controlled trial (RCT) of this. The design was for a three-armed RCT that investigated (i) a partial intervention that fundamentally relied on a brief programme of professional development (PD) and exemplary lessons to illustrate the approach that were taught spaced out during the year and (ii) a full intervention that in addition to these elements included a programme of modified lesson study phased over the teaching period October (2021) – March (2022).

Context

The context of the Centres for Excellence in Mathematics is a response to government concern about the mathematical skills of adults, particularly with respect to skills needed for work. This has long been a concern identified by various stakeholders (e.g., Confederation of British Industry, 2015). The importance of mathematics was highlighted in the current government's Industrial Strategy (BEIS, 2017), with indications of an intention to address deficits where they occur. Professor Sir Adrian Smith's review of post-16 mathematics in 2017 (Smith, 2017) considered ways of improving mathematics education for the 16-18 age cohort. This included a review of the mathematical needs of post-16 students and concluded that there was strong demand for mathematical and quantitative skills in the labour market at all levels and consistent undersupply and adults with basic numeracy skills earn higher wages and are more likely to be employed than those who fail to master basic quantitative skills. The review also suggested that there was a need for further investment to improve mathematics teaching in Further Education (FE) colleges.

Over one third (34%) of students aged 16-18 study in either General FE or Sixth Form colleges (Association of Colleges, 2020), with the majority following vocational study programmes. Mathematics progress measures suggest that many students with low prior attainment in mathematics (i.e., GCSE grade 3 or below) do not make any measurable progress, in terms of their GCSE mathematics grade, by age 19 and GCSE pass rates for these students indicate that less than twenty percent attain the desired grade 4 or above (DfE, 2019). These data signal the scale of the problem.

The Wolf report (Wolf, 2011) on vocational education highlighted the need for mathematics skills but questioned the value of mathematics qualifications such as key skills and functional skills. The claim that GCSE was the 'gold standard' and that all students should have the opportunity to achieve this qualification triggered a series of changes to policy and curriculum, especially affecting post-16 students with low GCSE grades in mathematics.

Following the Wolf report the Condition of Funding (commonly referred to as the GCSE resit policy) was introduced from September 2014, which required all 16–18 year olds who had not already achieved GCSE grade C (replaced with grade 4 in a reform of GCSEs in 2017) to take a mathematics course as part of their study programme. The aim was for students to retake the GCSE examination but other 'stepping stone' qualifications such as Functional Skills Mathematics could be taken first if appropriate. The following year, from September 2016, it became compulsory for those with grade D (now grade 3) to take GCSE mathematics rather than a functional skills qualification, although the 'stepping stone' option was still available to those with lower grades. Subsequently, GCSE has become the main qualification taken by the majority of 16-18 year olds with low attainment in mathematics. Since the introduction of the Condition of Funding, there has been a review of, and subsequent changes, to functional skills qualifications and assessments.

The Condition of Funding has resulted in a large increase in the number of students taking mathematics qualifications at Level 2 and below, especially in General Further Education Colleges. There are few large-scale studies of mathematics teaching in the context of Further Education. Notably, a study that was nearing completion was led by CRME but was impacted by the outbreak of Covid-19 and has yet to report[1]. Studies that have explored various aspects of mathematics in the FE Context include for ETF (The Research Base, 2014), for the Department for Education (Higton et al., 2017), for NIACE (Robey & Jones, 2015) and the Nuffield Foundation (e.g., Noyes & Dalby, 2020) have explored various aspects of mathematics in the FE context. These studies suggest that mathematics learning for post16 students retaking mathematics needs to be a different experience to school and that alternative approaches and strategies are valued by students, many of whom have insecure understanding of basic concepts. Many of these studies identify a lack of motivation amongst students, which is often linked to negative prior experiences of mathematics and the emotional effects of failure following the GCSE examination. Students

[1] See Maths-for Life at https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/maths-for-life

are often reluctant to re-engage with mathematics, which may be evidenced by poor attendance or other avoidance behaviours, such as a lack of effort even when present in lessons (Dalby, 2014). Motivation is often gained from the 'exchange value' of the qualification rather than the 'use-value' of the mathematics skills being learned (Noyes & Dalby, 2020; Williams, 2012).

Teaching for Mastery

The National Centre for Excellence in Teaching Mathematics (NCETM) has, for some time now, been focussing on an approach to teaching mathematics in schools (first in Primary schools and more recently in Secondary schools) that embraces practices that have been observed in mathematics teaching and learning in the Far East particularly in Shanghai, China. The motivation to looking to this jurisdiction is that of the achievement of their students in international studies such as the OECD's PISA study. In general, they have higher levels of achievement and less of a tail of underachievement (Coles and Helme, 2022). Consequently, much of recent work of NCETM has focused on a model of 'teaching for mastery'[2] that has been very much informed by teaching in Shanghai, China, but developed with knowledge of teaching in the UK. This was initially informed by an exchange programme between Shanghai and English Primary teachers. A number of longitudinal evaluations have been reported by researchers at Sheffield Hallam University[3], but there has been no robust evidence of impact on learner outcomes. As Boylan and colleagues point out Teaching for Mastery can be considered government policy although not explicitly made as a clear policy statement (Boylan et al, 2018).

Following the pilot trials and the hiatus in research caused by the Covid-19 pandemic the research team in discussions with DfE chose to implement an RCT in Teaching for Mastery in FE. However, the decision was made to continue to work with key principles that had been developed by teachers and other stakeholders in FE, and especially the Centre Leads of CfEM. These key principles were derived with a particular set of beliefs and values relating to the students who undertake resit courses in FE: in particular, issues of motivation. The five key principles are set out in a Handbook that was available to all teachers including all involved in any of the three arms of the trials. The discussion that informed the development of these key principles was informed by a discussion paper prepared by the UoN CRME research team.

The five key principles for teaching in FE are:

1. Teaching that allows students to develop an understanding of mathematical structure. Teachers use representations to support understanding so that students know the 'why' and not just the 'how'. Representations can both clarify the meaning of a concept and provide access to the structure of mathematical problems.

[2] For current manifestations of the NCETM approach to Teaching for Mastery see https://www.ncetm.org.uk/teaching-for-mastery/
[3] See for example, Boylan et al. (2018), and Boylan et. Al (2019)

2. Valuing and building on students' prior learning. Teachers can celebrate and build on what students already know and make maximum use of the available teaching time to fill in key gaps in knowledge and understanding.
3. Prioritising curriculum coherence and connections. Students need to be encouraged to see the links between mathematical concepts e.g., similarity, ratio, trigonometry rather than seeing them as disparate entities that need to be individually learned. As well as these curricular links, using familiar representations across different topics will develop flexibility in their use, and support problem solving.
4. Developing both fluency and understanding of key ideas. Covering key content in depth to attain fluency and understanding that can be applied in different contexts is preferred to superficial coverage of a larger amount of material. It is important to remember that fluency is not just about knowing facts and procedures, but also how and when to use them.
5. Developing a culture in which everyone believes everyone can succeed. Embedding the culture that effort leads to improvement through low threat/high challenge tasks and activities. Success is of course relative; for some students it may be passing over the Grade 4 GCSE other threshold whilst for others in their first year of post-16 it may be starting this journey by moving from Grade 1 to Grade 2.

Special characteristics of the cohort

There are reasons to believe that the cohort of students that were participating in the RCT were not typical when compared to students pre-covid. As above, we identify two issues that may have impacted on the trial:

1. The students had a disrupted two years of education leading to the GCSE examination due to covid, including having to learn from home (if possible) at times;
2. The prior attainment levels of the cohort were likely to be less than would have been the case in 2019: for example, some students with Teacher Assessed Grade 3, may have obtained Grade 2 in previous GCSE examinations.

This latter point is detectable by comparing GCSE grades awarded over the last few years. Data of the 2019-2022 summer GCSE awards (England, by age) drawn from Joint Council for Qualifications[4] is used below to provide an insight into the potential unusual characteristics of the cohort of students.

- At age 16, the percentage of students getting grades U-2 during covid (2020-2021) is less than that before or after covid (2019 or 2022, respectively). On the other hand, the percentage of students getting grades 4+ during covid is greater than that before or after covid. The smaller U-2 percentage is not consistent with the National Reference Test finding that covid has negative impacts on secondary students' learning and achievement in mathematics.

[4] https://www.jcq.org.uk/examination-results

Table 3 Age 16 GCSE Mathematics cumulative percentages at key grade boundaries

Year	4+	3	U-2
2022	75.1%	10.5%	14.4%
2021	77.9%	9.9%	12.2%
2020	77.1%	11.1%	11.8%
2019	71.5%	12.7%	15.8%

- Looking at the age 17+ data, the percentage of students getting grades U-2 during covid (only 2021 data is available; 2020 data that is grouped by age and England is not available) is less than that before covid (2019). On the other hand, the percentage of students getting grades 3+/4+ during covid is greater than that before covid.

Table 4 Age 17+ GCSE Mathematics cumulative percentages at key grade boundaries

Year	4+	3	U-2
2022	20.1%	27.3%	52.6%
2021	38.6%	28.2%	33.2%
2020	NA	NA	NA
2019	21.2%	37.5%	41.3%

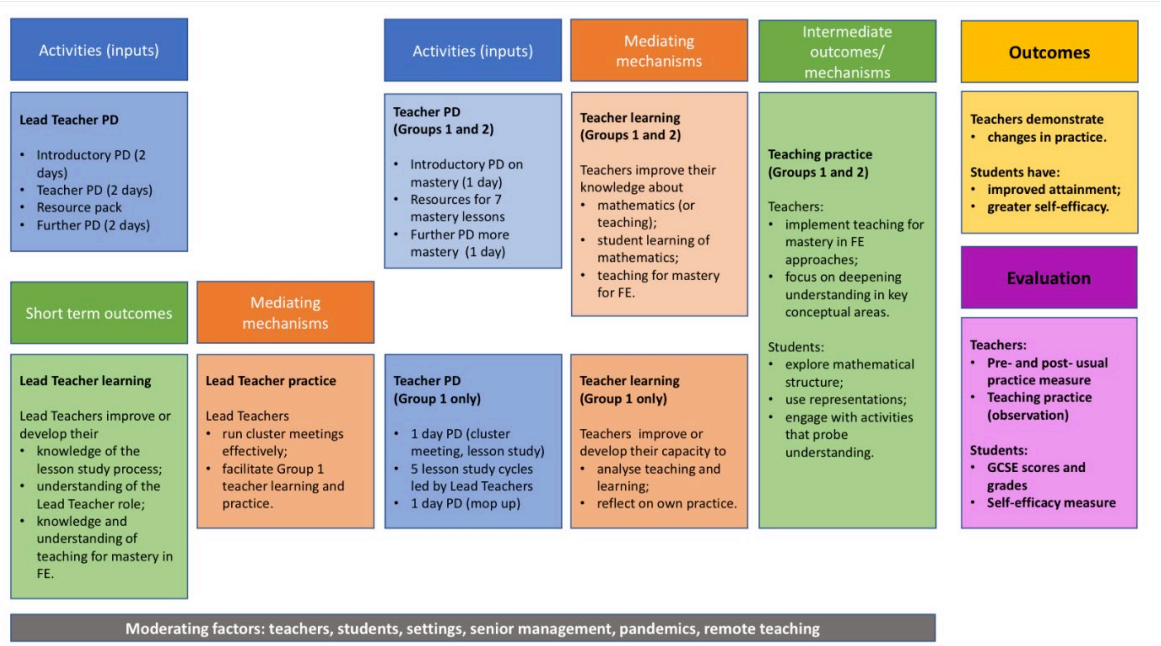
These data all suggest that the cohort of students participating in the RCT was probably not entirely comparable with a similar cohort from 2019 or earlier. Of course, it is not possible to determine the likely impact that these issues would have on the trials, and we should bear in mind the randomisation process that should ensure comparability across the different arms of the trials.

Brief name

CfEM Teaching for Mastery (TfM)

Why (rationale/theory)

Figure 1 Logic model of the project



Who

This randomised control trial involved an intervention, for teachers of GCSE resit classes, consisting of professional development and materials to support adopting a 'teaching for mastery' approach.

Three groups of teachers took part, those who received

- The full intervention (three full professional development days, five local cluster meetings led by a lead teacher, and five sets of teaching materials)
- A partial intervention (three full professional development days and five sets of materials)
- No intervention.

What (materials)

All intervention teachers were provided with

- A Teaching for Mastery handbook
- Detailed lesson plans and supporting resources for five Teaching for Mastery in FE Maths lessons.

Full-intervention teachers were further provided with online professional development materials focussed on familiarisation with the design of the lessons. These were particularly intended for those who missed cluster meetings.

The design of the lesson resources and more details about the content of the lessons can be found in Appendix A.

Lead teachers were provided with all the materials given to teachers in the full intervention group and detailed support materials to run the cluster meetings.

What (procedures)

Table 5, below, summarises all the trial teacher and lead teacher activity for the trial. More details about the aims of the professional development and the content of the sessions can be found in Appendix B.

Table 5 Summary of all Trial Teacher and Lead Teacher activities

Date	Activity
24th March 2021 Online	Lead teachers participate in a professional development day, led by the University of Nottingham.
23rd June 2021 Online	Lead teachers participate in a professional development day, led by the University of Nottingham.
6th and 8th October 2021 *Mornings*, Online	Lead teachers and trial teachers (both the full intervention and the partial intervention groups) participate in professional development on teaching for mastery, led by the University of Nottingham.
6th and 8th October 2021 *Afternoons*, Online	Lead teachers and trial teachers (the full intervention group) participate in professional development on lesson study and cluster meetings (two sessions available with each participant to attend one), led by the University of Nottingham.
13th and 15th October 2021 *Mornings*, Online	Lead teachers and trial teachers (both groups) participate in professional development on teaching for mastery (two sessions available with each participant to attend one), led by the University of Nottingham[5].
15th October 2021 *Afternoon*, Online	Lead teachers and the full intervention group participate in Cluster Meeting 0, to prepare for teaching Lesson 1 and Cluster Meeting 1, led by lead teachers.
Autumn term 2021	Trial teachers (both groups) teach Lessons 1 and 2 in Windows 1 and 2. Window 1: 1st to 26th November 2021 Window 2: 29th November to 17th December 2021 Teachers in the full-intervention group participate in Cluster Meetings 1 and 2 (the full intervention group), led by lead teachers.
5th January 2022	Lead teachers participate in a professional development day, led by the University of Nottingham.
12th and 14th January 2022	Lead teachers and trial teachers (both groups) participate in professional development on teaching for mastery, (two sessions available with each participant to attend one), led by the University of Nottingham.

[5] Two online twilight sessions were offered on 2nd and 4th November 2021 as a 'catch-up' for Trial Teachers who were unable to join these sessions.

Spring term 2022	Trial teachers (both groups) teach Lessons 3, 4, and 5 in Windows 3, 4 and 5. Window 3: 6th to 28th January 2022 Window 4: 31st January to 4th March 2022 Window 5: 7th March to 1st April 2022. Teachers in the full-intervention group participate in Cluster Meetings 3 (online), 4 (online) and 5, led by lead teachers.
22nd and 24th June	Trial teachers (both groups) participate in post-intervention professional development event (two sessions to choose from), led by the University of Nottingham.
6th July 2022	Lead teachers participate in a post-intervention PD event, led by the University of Nottingham.

Note: events were held face-to-face unless stated otherwise

Who (implementers)

The intervention was implemented by the University of Nottingham and ten Lead Teachers. The University of Nottingham provided professional development for the Lead Teachers, which prepared them for their role in providing support for Trial Teachers in the full intervention group.

Where (setting)

The Lead Teacher (LT) professional development days took place online. The final LT professional development event (post-intervention) took place at the University of Nottingham.

The Trial Teacher (TT) professional development days took place online. The final TT professional development events (post-intervention) took place at the University of Nottingham and in London (Friends' Meeting House).

The cluster meetings took place in colleges unless stated that they were held online (to take account of Covid-19 concerns).

CfEM trial lessons were taught in students' normal classrooms for mathematics (equipped with a data projector to work with the lesson introductory materials).

When and how much (dosage)

Teachers were asked to teach Lessons A and B at any time and Lessons 1 to 5 within specific windows to all their GCSE classes. They were encouraged to adopt Teaching for Mastery approaches, as exemplified in the sample lessons, to all their classes including in other lessons.

Tailoring

In the professional development events in October and January, the nature and extent of changes teachers could make to the lessons were explicitly discussed. Lead Teachers also addressed the issue in cluster meetings of full intervention teachers. Teachers were encouraged to adapt the lessons for their own classes.

Evaluation objectives

Impact evaluation

The Teaching for Mastery in FE trial evaluation sought to address the following research questions (RQs). These were set out in the evaluation protocol, discussed further in the statistical analysis plan.

The evaluation addressed the following research questions:

Primary research question

1. Does the Mastery for FE programme improve 16-19 further education students' overall GCSE mathematics resit performance when measured by the GCSE raw point scores?

Secondary research questions

2. Does the Mastery for FE programme increase the probability of students achieving a grade 4 on their GCSE mathematics resit?
3. Does the Mastery for FE programme increase the probability of a student scoring a level higher than previously achieved on their GCSE mathematics exam?
4. Does the Mastery for FE programme improve students' mathematical self-efficacy?

Implementation and process evaluation (IPE)

The IPE addressed the following research questions:

1. **Fidelity:** To what extent do implementers adhere to the intended model? In particular:
    - 1.1. To what extent does the professional development provided for the Lead Teachers adhere to the intended model?
    - 1.2. To what extent do Lead Teachers implement the cluster meetings in line with the model in the resources? (The full intervention group)
    - 1.3. To what extent are the objectives of the planned professional development activities met? (The full intervention group)
    - 1.4. To what extent are the objectives of the planned professional development (mastery) activities met when implemented? (The full intervention and the partial intervention groups)
    - 1.5. To what extent do Trial Teachers adhere to the mastery approach in the sample lessons? (The full intervention and the partial intervention groups)
    - 1.6. What factors contribute to variation in fidelity?
2. **Dosage:** How much of the intervention is delivered and received? In particular:
    - 2.1. To what extent do Lead Teachers experience the recommended amount of PD?

2.2. To what extent do Trial Teachers experience the recommended amount of PD?

2.3. To what extent do students experience the recommended number of CfEM sample lessons?

2.4. What factors contribute to any variation in dosage?

3. **Responsiveness:** To what extent do participants engage with the intervention? In particular:

3.1. To what extent, and how, do Lead Teachers engage in PD activities?

3.2. To what extent, and how, do the full intervention Trial Teachers engage in PD activities?

3.3. To what extent, and how, do the partial intervention Trial Teachers engage in PD activities?

3.4. To what extent do Trial Teachers apply the mastery approaches exemplified in the sample lessons and does this change over time?

3.5. To what extent do students engage in CfEM lessons and does this change over time?

4. **Programme differentiation:** To what extent is the intervention distinguishable from existing classroom practice and/or accessed professional development? In particular:

4.1. Is the Trial Teachers' implementation of the CfEM mastery approaches significantly different from their typical practice? If so, how, and does this change over time?

4.2. Have Trial Teachers (the partial intervention and the control groups) received non-CfEM PD similar to that received by the full intervention group (lesson study), either prior to, or during, the intervention?

4.3. Have Trial Teachers (the control group) received PD of a similar nature to the full intervention and the partial intervention groups (mastery)?

5. **Quality:** How well is the intervention implemented?

5.1. How well does the PD prepare and support Lead Teachers to facilitate cluster meetings?

5.2. How well do cluster meetings encourage Trial Teachers to analyse teaching and learning and reflect on practice?

5.3. How well does the PD develop understanding of teaching for mastery?

5.4. How well do teachers practise teaching for mastery?

5.5. How well do students explore mathematical structure, use representations and engage with activities that probe understanding in the mastery lessons?

5.6. What factors contribute to variation in implementation quality?

6. **Mediating mechanisms:** Are the hypothesised mediating mechanisms present? In particular:

6.1. Are the hypothesised mediating mechanisms that arise from the PD for both Lead Teachers and Trial Teachers present?

6.2. Are the hypothesised mediating mechanisms that arise from the CfEM lessons present?

6.3. Are there alternative or complementary mediating mechanisms at play?

Ethics and trial registration

The study was reviewed and approved by the University of Nottingham School of Education Ethics Committee on 16th March 2021. The trial was registered with the Open Science Foundation (osf.io/gtx8u) prior to carrying out the impact analysis in January 2023.

Data protection

A Memorandum of Understanding (MoU) was signed with each teacher and Further Education college, as well as a data sharing agreement which set out the aims and benefits of the data share, the agreed purposes, personal and special category data to be collected, legal basis for processing, as well as data handling and data security measures to be employed.

The agreed legal basis for processing of personal data was Article 6(1e): "Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller". For special category data, the legal basis was Article 9(2j): "processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes".

As set out in the privacy notice, the project collected data under the terms of the University's Royal Charter in our capacity as a teaching and research body to advance education and learning, and specifically to evaluate the impact of the Mathematics Mastery in FE intervention using a robust randomised controlled trial design. Surveys and implementation and process data was processed under Article 6(1a) "the data subject has given consent to the processing of his or her personal data for one or more specific purposes." Copies of the documentation provided to students, teachers, and FE colleges can be found in the appendices.

Project team

The University of Nottingham team comprised the following staff members:

- Professor Geoff Wake, Principal Investigator, and Professor of Mathematics Education at the University of Nottingham, overseeing the team, evaluation, stakeholder relations, reporting, and dissemination.
- Professor Andrew Noyes, Co-Principal Investigator, and Professor of Education at the University of Nottingham.
- Dr Diane Dalby, Senior Research Fellow collected data for the implementation and process evaluation and assisted with report writing.
- Dr Marie Joubert, Senior Research Fellow led the professional development for Lead Teachers and Trial Teachers, the development of the lessons, disseminated on aspects of emerging research, collected data for the implementation and process evaluation and assisted with report writing.
- Dr Marc North, Assistant Professor in Education (between September 2019 and August 2020) contributed to the design and piloting of the teacher and student self-efficacy surveys, CfEM lesson resources (for the Contextualisation, Mastery, and Technology and Data themes), and data collection via observations and interviews to gauge the efficacy of the lesson resources.

- Dr Michael Adkins, Senior Research Fellow, led the impact evaluation carrying out the data analysis and assisted with report writing.
- Dr Gabriel Lee, Senior Research Fellow, led the quantitative analysis of the implementation and process evaluation and assisted with report writing.
- Dr Jonathan Halls, Assistant Professor (September 2021 to April 2022) contributed to case study data collection for the Implementation and Process Evaluation, including classroom observations, teacher interviews and summarising individual cases.
- Kanchana Minson, CfEM Project Manager (between July 2019 and June 2021) supported the UoN project team establishing the governance structure for the project and liaising with the funder, delivery partners and advisory group. She managed the delivery of the project through all aspects of project management through the early part of the project lifecycle.
- Alex Phillips, CfEM Project Manager (between May 2021-Jan 2023) supported the UoN project team being responsible for the day-to-day management of the project, co-ordinating with the funder, delivery partners and advisory group, and was critical to the design of the project's data management and collection processes.
- James Fox, Project Officer, joined the team in March 2021 and his key areas of responsibility included project delivery and supporting the research team on data collection, stakeholder engagement and event planning.

Methods

Impact evaluation

Trial design

The trial was designed as an efficacy study for approximately 8500 students in 130 Further Education College (FE) settings. The intervention was created to improve GCSE Mathematic retake outcomes for post-16 students with a focus on the Further Education Colleges, but applicable to those more widely within 6th form colleges, schools and other training providers. Students between the ages of 16-20 undertaking GCSE Mathematics resits as part of the condition of funding requirement to continue to study Mathematics as part of their post-16 study programme with the aim of achieving a grade 4 “pass”.

At the design stage of the trial, the aim was to recruit students from 130 FE settings, with an average cluster size of 65 students. FE settings were recruited from across the 21 Centres for Excellence in Mathematics networks via an initial expression of interest. From these, 147 settings went forward for randomisation in July 2021 and were assigned to a full intervention arm consisting of materials and lesson study, a partial intervention arm consisting of materials only, and a business-as-usual control. The trial design parameters are summarised in Table 6.

Table 6 Trial design

Trial Design Parameters	Notes
Trial design	Three arm cluster randomized
Unit of randomisation	Further Education Setting
Primary outcome	GCSE z-score adjusted raw score, collected directly from FE Colleges.
Secondary outcome(s)	Mathematics Mastery subscale z-score derived from GCSE Mathematics item-level data collected directly from FE Colleges. GCSE Mathematics Grade Score, collected directly from FE Colleges.
Baseline for primary outcome	KS2 Mathematics z-score, collected from the National Pupil Database and equated between two measures.
Baseline for secondary outcome	KS2 Mathematics z-score, collected from the National Pupil Database and equated between two measures.

Participant selection

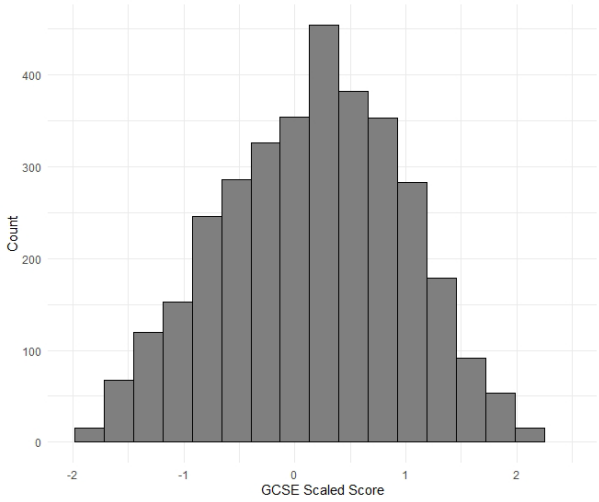
Students in participating settings were eligible if they had not achieved a grade 4 “pass” in GCSE Mathematics at age 16, were taking GCSE Mathematics resit as part of the condition of funding and were aged between 16-20 (by June 2022). Colleges were required, once all students were assigned to classes within FE Colleges, to provide critical data points for each student to allow for matching to the DfE’s National Pupil Database – for example the Unique Learner Number (ULN), forenames, surname, date of birth, and gender. Later in the trial FE Colleges were required to provide a list of participating students that had achieved a grade 4 in the November 2021 resit, and the exam identifier for each student assigned by the exam board.

Outcome measures

Primary outcome measures

As the trial was targeted at students undertaking GCSE Mathematics resits as part of the 16-19 Condition of Funding policy, all participating students were sitting the foundation paper as part of their studies. This limited the grade 9 – 0 scale to 5 – 0. In discussion with the Centres for Excellence partners, the expectation was that the majority of these retakes involved those students that had previously achieved a grade 3 at Key Stage 4 taken at age 16, again limiting discrimination between the students. Therefore, this trial used the GCSE raw score as reported by exam boards, and collected directly from participating FE colleges, and shared via a data sharing agreement with the University of Nottingham. Differences between four exam boards – Assessment and Qualifications Alliance (AQA), Pearson Edexcel, Oxford Cambridge Recognition (OCR), and WJEC needed to be equated. To put these on the same scale, the total score from all papers were standardised separately to have means of 0 and standard deviations of 1 to create a single scale. The distribution is presented in Figure 2.

Figure 2 Distribution of GCSE Mathematics post-intervention scaled score (Source: ONS)

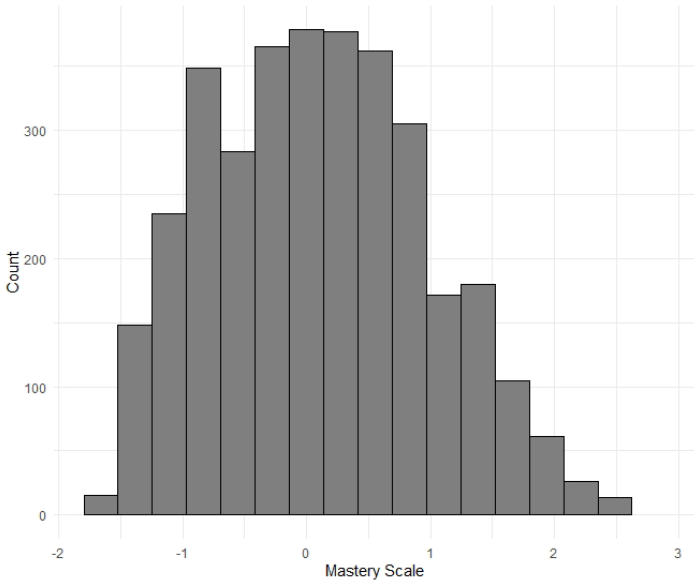


Secondary outcome measures

A Mastery based subscale to focus on the examination questions related directly to the mathematical content was also created from the question item-level scores coded independently by three University of Nottingham mathematics education specialists (the questions identified are presented in Appendix C).

Agreement in coding was high and any disagreements were resolved by discussion. Scores were created by summing and standardised by exam board. The distribution is presented in Figure 3.

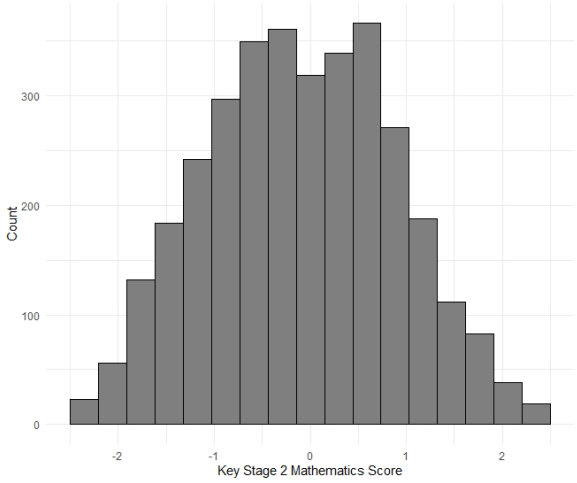
Figure 3 Distribution of GCSE Mastery scale (Source: ONS)



Baseline measures

The baseline measure across the primary, secondary and subgroup analyses was the Key Stage 2 national test score in Mathematics sat by age 11 students in all state funded primary schools in England. This was drawn through the data share with the DfE's National Pupil Database (NPD) archive. As an amendment to the planned analysis outlined in the Statistical Analysis Plan, due to the DfE's condition of funding policy requiring multiple retakes until the age of 18, the participants were drawn from a number of different KS4 cohorts that fell either side of a minor change in scale of the KS2 Mathematics score. To resolve this, we standardised the two measures to put these on the same scale. Figure 4 presents a histogram illustrating the distribution of the standardised KS2 Mathematics score for all participating students.

Figure 4 Distribution of KS2 Mathematics pre-intervention scaled score (Source: ONS)



Calculation of sample size

As discussed in the protocol, we used the R package *PowerUpR* (Bulus et al., 2021) to calculate the appropriate sample size for the trial. The trial was designed as a three-arm trial comprising of the full

intervention group vs. the control group, with an additional partial intervention group would be underpowered compared to the control group. At the design stage, we made the following assumptions:

- α = 0.05
- β = 0.8
- Two-sided hypothesis.
- An estimate of 65 students per college.
- An Intra-Class Correlation Coefficient (ICC) of 0.2 at the College level based on the EEF trial of the Maths 4 Life in which the University of Nottingham acted as the developer.
- Baseline measure is the fine graded KS2 mathematics score have correlation 0.55 (and therefore a R^2 of 0.30) with our KS5 mathematics GCSE outcome.
- Baseline resit pass-rate: 28% based on prior assumptions from the EEF funded Maths 4 Life trial.

At the design stage, for the main effect we estimated that a MDES of 0.239 was achievable, along with an MDES of 0.277 for the partial intervention vs. control comparison. As an example, subgroup analyses such as those students previously in receipt of free school meals within 6 years was underpowered, but with the proposed samples in each group, an MDES of 0.252 and 0.292 was still achievable.

Randomisation

FE settings were randomised in one batch on 19th July 2021 using simple randomisation approach using the R package *randomizr* (Coppock et al., 2023). Probabilities assigned to each condition were as follows 0.384 for the full intervention and control, and 0.232 for the partial intervention. Balance was checked against three variables – the average number of students reported by participating colleges, the average number of mathematics groups reported by colleges and the average progress in mathematics reported in 2019.

Statistical analyses

Primary efficacy analysis

We investigated the impact of Maths Mastery in FE on mathematics achievement on the score achieved in the summer 2022 GCSE mathematics by a standardised GCSE raw score. This was adjusted for prior attainment using KS2 Mathematics score. Finally, randomisation took place at the FE setting-level, and so the score the GCSE Mathematics exam was modelled using a multilevel varying intercept model which is presented below:

$$y_{ij} = \beta_{0j} + \beta_1 Treatment_{ij} + \beta_2 KS2_{ij} + u_{0j} + \varepsilon_{ij}$$
$$u_{0j} \sim N\left(0, \sigma^2_{FESetting}\right)$$
$$\varepsilon_{ij} \sim N\left(0, \sigma^2_y\right)$$

This equation can be understood as follows. The post-test score for the *i*th student in the *j*th FE college setting is equal to the grand mean score (β_{0j}), plus the impact of a binary indicator denoting treatment received (β_1) which is coded as 0 or 1, plus the impact of the mean-centred normally distributed KS2 Mathematics attainment measure (β_2), plus the college setting-level error term (u_{0j}), and finally plus the student-level error term (ϵ_{ij}). The two error terms each receive their own probability distribution which are

assumed to be normally distributed and centred on 0, with the two variance parameters which were estimated from the data ($\sigma^2_{FE\ Setting}$ and σ^2_y).

Secondary analysis

Three secondary efficacy analyses were analysed: the impact of the Maths Mastery materials on mathematics achievement measured by the standardized GCSE Mathematics raw score; the impact of the intervention as measured by the GCSE grade score; and lastly the impact of the mastery item GCSE subscale. These were modelled in same manner as the primary analysis discussed, with the same prior distributions, on the basis of intention to treat.

Compliance to the mastery in FE intervention was simplified from that planned within the protocol and statistical analysis plan to focus on whether the lessons had been delivered across all trial teacher classes and was assessed at the settings-level. An instrumental variables approach was followed, although it was again simplified due to software limitations within the Secure Research Environment and fitted as a single level two-stage least squares model:

$$compliance_j = \beta_0 + \beta_1 Treatment_i + \beta_2 Pre - Test_i + e_{ij}$$
$$y_{ij} = \beta_{0j} + \beta_1 \widehat{Compliance}_j + \beta_2 Pre - test_{ij} + \epsilon_{ij}$$
$$\epsilon_{ij} \sim \mathcal{N}(0, \sigma_y^2)$$

Missing data analysis

We carried out descriptive investigation of missingness on the primary outcome, covariates and potential auxiliary variables. A multilevel logistic regression model of missingness was then fitted to understand the probability of missingness for the outcome and pre-test using the R package *lme4* (Bates et al., 2015).

A two-level multilevel multiple imputation model was carried out using the primary model's outcome, covariates, FE settings index and two auxiliary variables – KS4 academic year, and KS4 Attainment 8 score. The imputation was carried out using the joint modelling package *Jomo* in R (Quartagno & Carpenter, 2022). The sampler was run for 100,000 iterations, generating four imputed datasets.

Visual checks of the Markov Chain Monte Carlo chain traces and potential scale reduction factors (Rhat) were assessed for evidence of convergence. All parameters had a rhat score of <1.1 indicating good convergence. Further checks on the Auto-Correlation Function (ACF) statistic indicated no issues with autocorrelation of the samples. The four imputed datasets were extracted, and the primary and FSM model were fitted, and posterior distributions combined. Summary statistics were generated, along with effect sizes.

Sub-group analyses

Four sub-group analyses were carried out on the primary and Maths mastery subscale outcome discussed above for the full intervention vs. control, and the partial intervention vs. control contrasts. These were be modelled using an interaction model to investigate the impact of the intervention for disadvantaged students using the free school meals in the past 6 years flag from the National Pupil Database for students

when they were 16, as Further Education Colleges do not collect this data post-16. The model is presented below:

$$y_{ij} = \beta_{0j} + \beta_1 Treatment_{ij} + \beta_2 KS2_{ij} + \beta_3 FSM6_{ij} + \beta_4 Treatment_{ij} * FSM6_{ij} + u_{0j} + \epsilon_{ij}$$
$$u_{0j} \sim N\left(0, \sigma^2_{FESetting}\right)$$
$$\varepsilon_{ij} \sim N\left(0, \sigma^2_y\right)$$

Estimation of effect sizes

Effect sizes were estimated by extracting the posterior simulations from our fitted models, and as discussed in Wake et al. (2022) this provides the benefit of computing direct probabilities of effect sizes, posterior indices such as the probability of direction – the probability of a parameter being strictly positive or negative (Makowski et al, 2019), and estimating the proportion of the interval falling within a "region of practical equivalence" or ROPE (see Kruschke, 2018, p. 270). The latter compares where the range of credible values fall on a 95% highest density interval in relation to a fixed width of an effect size – in our case ±0.1. If the interval falls entirely within the effect size range, a null result can be accepted – i.e., no practical significance to the result, if the interval falls entirely outside the effect size range then there is strong evidence of practical significance, and if the interval falls within and outside the effect size range then the evidence is mixed.

Estimation of ICC

Intra-Class Correlation (ICC) was calculated for the primary and secondary outcomes to assess the variability at the FE setting level, and to assess the plausibility of the assumptions behind the power calculations performed at the design stage. These were estimated using a multilevel variance components model – an empty regression model with the varying intercept settings adjustments only, but no covariates such as the treatment condition and pre-intervention attainment score. This was estimated by Bayesian inference using weakly informative priors, and calculated using the following equation:

$$ICC = \frac{\sigma^2_{FE\ Setting}}{(\sigma^2_{FE\ Setting} + \sigma^2_y)}$$

Longitudinal analysis

No follow-up longitudinal analyses were planned or commissioned, and in agreement with the Department for Education, the data does not have permission to be archived.

Implementation and process evaluation

Recall the logic model was depicted in Figure 1. Data was collected using the methods described below. Qualitative case studies were carried out with a sample of settings in the intervention groups (see following section for sampling strategy). Table 7 summarises the outcome measures and methods that were used to answer the research questions.

Table 7 IPE questions, data collected and instruments used

Research Question	Outcome measure	Method
1. Fidelity: To what extent do implementers adhere to the intended model?		
1.1 To what extent does the professional development provided for the Lead Teachers adhere to the intended model?	Lead Teacher trainers' adherence to training plan	Lead Teacher trainers' reflective log
1.2 To what extent do Lead Teachers implement the cluster meetings in line with the planned model in the resources? (The full intervention group)	Lead Teacher adherence during PD sessions - level of adherence to session plan	PD session evaluations
1.3 To what extent are the objectives of the planned professional development activities met? (The full intervention group)	Adherence to PD plans Adherence to cluster meeting plans	PD session evaluation (including cluster meetings) Lead Teacher interviews
1.4 To what extent are the objectives of the planned professional development (mastery) activities met when implemented? (The full intervention and the partial intervention groups)	Adherence to PD plans	PD session evaluation Trial Teacher interviews
1.5 To what extent do Trial Teachers adhere to the mastery approach in the sample lessons? (The full intervention and the partial intervention groups)	Adherence to lesson plans. Use of mastery principles.	Lesson observations Trial Teachers interviews

1.6 What factors contribute to variation in fidelity?	Level of understanding demonstrated by Trial Teachers during periods of non-adherence/adaptation Level of understanding demonstrated by students during lessons during periods of non-adherence/adaptation Trial Teacher justifications for non-adherence/adaptation	Lesson observations Trial Teachers interviews

2. Dosage: How much of the intervention is delivered and received?		
2.1 To what extent do Lead Teachers experience the recommended amount of PD?	Number of sessions delivered to Lead Teachers Length of each session delivered Lead Teacher attendance at each session Lead teacher attendance at Teacher PD sessions	Lead Teacher PD session logs
2.2 To what extent do Trial Teachers experience the recommended amount of PD?	Number of sessions delivered Length of each session delivered Trial Teacher attendance at each session	Lead Teacher PD session logs PD session logs
2.3 To what extent do students experience the recommended number of CfEM sample mastery lessons?	Number of sessions delivered by each Trial Teacher to each of their GCSE resit classes[6] Length of each lesson taught by Trial Teacher in	Trial Teacher logs

[6] This will be used for the compliance indicator in the analysis of treatment effects in the presence of non-compliance.

	each setting to each of their GCSE resit classes	
2.4 What factors contribute to any variation in dosage?	Trial Teacher perceptions of contributing factors.	Trial Teacher interviews

3. Responsiveness: To what extent do participants engage with the intervention?		
3.1 To what extent, and how, do Lead Teachers engage in PD activities?	Lead Teacher engagement during PD sessions Lead Teacher perceptions of personal and peer engagement in sessions	Lead Teacher PD session evaluations Lead Teacher interviews
3.2 To what extent, and how, do the full intervention Trial Teachers engage in PD activities?	Trial Teachers engagement during PD sessions Lead Teacher perceptions of personal and peer engagement in sessions	Trial Teacher PD session evaluations Trial Teacher interviews Trial Teacher interviews
3.3 To what extent, and how, do the partial intervention Trial Teachers engage in PD activities?	Trial Teacher engagement during PD sessions	PD session evaluations Trial Teacher interviews and survey
3.4 To what extent do Trial Teachers apply the mastery approaches exemplified in the sample lessons and does this change over time?	Teacher use of the lesson resources	Lesson observations Trial Teacher interviews and survey
3.5 To what extent do students engage in CfEM lessons and does this change over time?	Student engagement in CfEM lessons	Lesson observations Student interviews Class Teacher interviews

4. Programme differentiation: To what extent is the intervention distinguishable from existing practice?		
4.1 Is the Trial Teachers' implementation of the CfEM mastery approach significantly different from their typical practice? If so, how, and does this change over time?	Class Teacher perceptions of personal practice	Trial Teacher interviews and survey

4.2 Have Trial Teachers (the partial intervention and the control groups) received non-CfEM PD similar to that received by the full intervention group (lesson study), either prior to, or during, the intervention?	Teacher self-reported engagement in other PD	Trial Teacher interviews and survey
4.3 Have Trial Teachers (the control group) received PD of a similar nature to the full intervention and the partial intervention groups (mastery)?	Teacher self-reported PD participation	Trial Teacher interviews and survey

5. Quality: How well is the intervention implemented?		
5.1 How well does the professional development prepare and support the Lead Teachers to facilitate cluster meetings?	Lead Teacher engagement in PD	PD session evaluations Lead Teacher interviews
5.2 How well do cluster meetings encourage Trial Teachers to analyse teaching and learning and reflect on practice?	Trial Teacher behaviour in cluster meetings	Cluster meeting evaluations Trial Teacher interviews Lead Teacher interviews
5.3 How well does the PD develop understanding of teaching for mastery?	Trial Teacher perceptions of their own learning Lead Teacher observations of Teacher learning (Group 1)	Trial Teacher interviews and survey Lead Teacher interviews
5.4 How well do teachers practise teaching for mastery?	Trial Teacher perception of personal practice Teacher behaviour	Trial Teacher interviews and survey Lesson observations
5.5 How well do students explore mathematical structure, use representations and engage with activities that	Student behaviour in CfEM lessons	Student interviews Classroom observations

probe understanding in the mastery lessons?		
5.6 What factors contribute to variation in implementation quality?	Trial Teacher perceptions of contributing factors Student perceptions of contributing factors	Trial Teacher interviews Student interviews

6. Causal mechanisms: Are the hypothesised mediating mechanisms present?		
6.1 Are the hypothesised mediating mechanisms that arise from the PD for both Lead Teachers and Trial Teachers present?	Lead Teacher facilitation of cluster meetings Trial Teacher discussion during PD sessions Trial Teacher behaviour during lessons Student discussion and behaviour during lessons Trial Teacher perceptions of the nature of the intervention Student perceptions of the nature of the intervention	PD session evaluations Lesson observations Trial Teacher interviews and survey Student interviews
6.2 Are the hypothesised mediating mechanisms that arise from the CfEM lessons present?	Trial Teacher perceptions of the nature of the intervention Student perceptions of the nature of the intervention	Trial Teacher interviews and survey Student interviews
6.3 Are there alternative or complementary mediating mechanisms at play?	Trial Teacher behaviour during PD sessions Trial Teacher behaviour during lessons Student behaviour during lessons Trial Teacher perceptions of the nature of the intervention Student perceptions of the nature of the intervention	PD session evaluations Lesson observations Trial Teacher interviews and survey Student interviews

Administrative data

Dosage data was collected at three levels. The Project Leads from the University of Nottingham collected session logs that recorded the number and duration of PD sessions delivered to Lead Teachers and Trial Teachers (the full intervention and the partial intervention groups). Lead and Trial Teacher attendance rates

were be recorded for each session. Trial Teachers (the full intervention and the partial intervention groups) completed a lesson log to record the number of lessons delivered to each of their GCSE resit classes. The number of lessons delivered by each Trial Teacher was used for the compliance indicator.

Online surveys

Trial Teachers in all three groups completed a survey to provide baseline data (e.g., on self-efficacy, teacher practice). After the intervention, teachers in all three groups completed a similar survey. Questions related to professional development during the intervention period were included but was bespoke to each group. Further questions relating to the other research themes were included in the surveys for the full intervention and partial intervention groups, as appropriate.

Observations

We conducted selective, systematic observations (Angrosino & Mays de Perez, 2000, p. 677), to ensure we were capturing the key information to help us answer our research questions (Kawulich, 2005).

- For classroom observations, this entailed using the CfEM teaching for mastery '5 Key Principles' (which describe the desired teaching approach), alongside the relevant lesson plan to assess quality and fidelity respectively.
- Observations were used to assess the responsiveness of Trial Teachers (to the PD) and students (to lessons). Whilst engagement was not always be visible to the observer, it was partially inferred from indicators such as the level of attention shown during presentations and the level of questioning and on-topic discussion from and amongst participants. We triangulated these findings with interview data, and administrative data.
- Finally, observations supported our evaluation of the hypothesised causal mechanisms in both PD sessions and lessons. To do this we looked for discussion and behaviour that verified, contradicted, or added to, the hypothesised causal mechanisms in the logic model. For example, in the case of PD, did teachers have discussions that (implicitly or explicitly) revealed a deepening of their understanding of teaching for Mastery? In the case of lessons, did we observe, for example, students using representations to understand mathematical structures?

Interviews

Lead Teacher, Trial Teacher and student perceptions of fidelity, quality, responsiveness and causal mechanisms were sought through interviews to build a more complete picture. The Lead Teachers and Trial Teachers who were observed were interviewed after their observations. A small number of students were interviewed, in small groups, from each case study setting. For all interviews, semi-structured guides were used to address the relevant research questions (see Table 4 for more details on which research questions will be covered by each set of interviews).

Case studies

The case studies allowed us to gather in-depth qualitative insights across all themes of the IPE. The unit of analysis was the trial teacher. Trial teachers from the full intervention (FI) group were selected by first

choosing three of the area-based cluster groups, so that there was representation from northern, southern and central regions. Two teachers were then selected from each of these clusters. For the partial intervention (PI) and control groups, a randomised approach was taken but with some re-selection to ensure the final sample achieved a balance across regions and gender.

The cases synthesised data from a range of methods (see Table 8) to understand the case of implementation more holistically. The Trial Teacher case studies were coded with the teacher group and setting ID. Three sets of case studies were developed, each set comprising Trial Teachers from one of the three groups:

- **Full Intervention group.** For the case studies from the Full Intervention group (6), the study followed one Trial Teacher and combined observations of six lesson study cluster meetings and five CfEM lessons with interviews with the Lead Teacher, the Trial Teacher and students. The Trial Teacher was interviewed at the start of the intervention (after the initial professional development), twice during the intervention, once in Window 2 and once in Window 4) and at the end of the intervention. The Lead Teacher was interviewed after each Lead Teacher professional development session and at the start of the intervention (after the initial professional development), twice during the intervention, once in Window 2 and once in Window 4) and at the end of the intervention. Case studies FI92, FI136, FI14, FI157, FI67, FI60 and LTs A, B and C.
- **Partial Intervention group.** For the case studies from the Partial Intervention group (2), the study followed one Trial Teacher and combined observations of five CfEM lessons with interviews with the Trial Teacher and students. Case studies PI46, PI89.
- **Control group**. For the case studies from the control group (4), three business-as-usual lessons were observed and the teachers were interviewed. Case studies C10, C56, C6, C94.

All Lead Teachers and Trial Teachers were anonymised and in the results section they are referred to by number.

Table 8 below provides details of the instruments used, the group of recipients of the intervention each instrument applied to and which research questions each instrument addressed. Appendices provide examples of the different instruments.

Table 8 IPE data collection overview

Instrument	Group	Research questions addressed
PD session evaluations	Lead teachers	3.1, 5.1, 6.1, 6.3
Interviews	Lead teachers	1.3, 3.1, 5.1, 5.2, 5.3
PD session log	Lead teachers (all until randomisation, then case studies only)	2.1
Cluster meeting evaluations (observations)	Full intervention case study teachers	1.3, 5.2, 5.3
Cluster meeting log	Full intervention teachers	2.2
PD session log	Full and partial intervention teachers	2.2
Trial Teacher PD session evaluations	Full and partial intervention teachers	3.2, 3.3, 6.1, 6.3
Trial Teacher interviews	Full and partial intervention case study teachers	1.4, 1.5, 1.6, 2.4, 3.2, 3.4, 5.6, 3.3, 3.5, 4.1, 4.2, 4.3, 5.2, 5.3, 5.4, 6.1, 6.2, 6.3
Trial Teacher interviews	Control group case study teachers	4.2, 4.3, 5.5
Lesson observations	Full and partial intervention case study teachers	1.5, 3.4, 3.5, 5.4, 5.5, 6.2, 6.3
Lesson observations	Control group case study teachers	6.3, 5.5
Trial Teacher teaching logs	Full and partial intervention teachers	2.3
Trial Teacher survey	Full and partial intervention and control group	3.3, 3.4, 4.1, 4.2, 4.3, 5.3, 5.4, 6.1, 6.2, 6.3
Student interviews (focus groups)	Full and partial intervention case studies	3.5, 5.5, 5.6, 6.1, 6.2, 6.3
Student surveys	Full and partial intervention and control group	4.1, 5.4, 5.5, 5.6

Data

The data gathered, and used in this IPE report includes survey responses, case study observations and interviews, activity logs. Survey data is summarised in Table 9.

Table 9 Surveys and response rates

Survey	Group (Total number)	N Responses
PD session 1 evaluation survey (Lead Teachers)	Lead teachers (10 teachers)	10 teachers
PD session 2 evaluation survey (Lead Teachers)	Lead teachers (10 teachers)	10 teachers
PD session 3 evaluation survey (Lead Teachers)	Lead teachers (10 teachers)	7 teachers
PD lesson studies and cluster meeting session evaluation survey (Trial Teachers)	Full Intervention (41 teachers)	29 teachers
PD mastery 1 & 2 session evaluation (Trial Teachers)	Full & Partial Intervention (70 teachers)	38 teachers
PD mastery 3 session evaluation (Trial Teachers)	Full & Partial Intervention (70 teachers)	57 teachers
PD mastery 4 session evaluation (Trial Teachers)	Full & Partial intervention (70 teachers)	46 teachers
Pre-intervention teacher survey (Trial Teachers)	Full & Partial Intervention & Control (123 teachers)	107 teachers
Post-intervention teacher survey (Trial Teachers)	Full & Partial Intervention & Control (123 teachers)	91 teachers
Pre-intervention student survey (Students)	Full & Partial Intervention & Control (123 teachers)	1012 students
Post-intervention student survey (Students)	Full & Partial Intervention & Control (123 teachers)	1079 students

Table 10 Case study data

Setting ID	Group	Observations	Interviews
10	Control	3	3
56	Control	2	2
6	Control	3	3
94	Control	3	3
92	Full Intervention	5	4
136	Full Intervention	5	4
14	Full Intervention	4	3
157	Full Intervention	4	3
67	Full Intervention	4	3
60	Full Intervention	4	3
46	Partial	4	4
89	Partial	5	4
38	Partial	1	1
		47	40

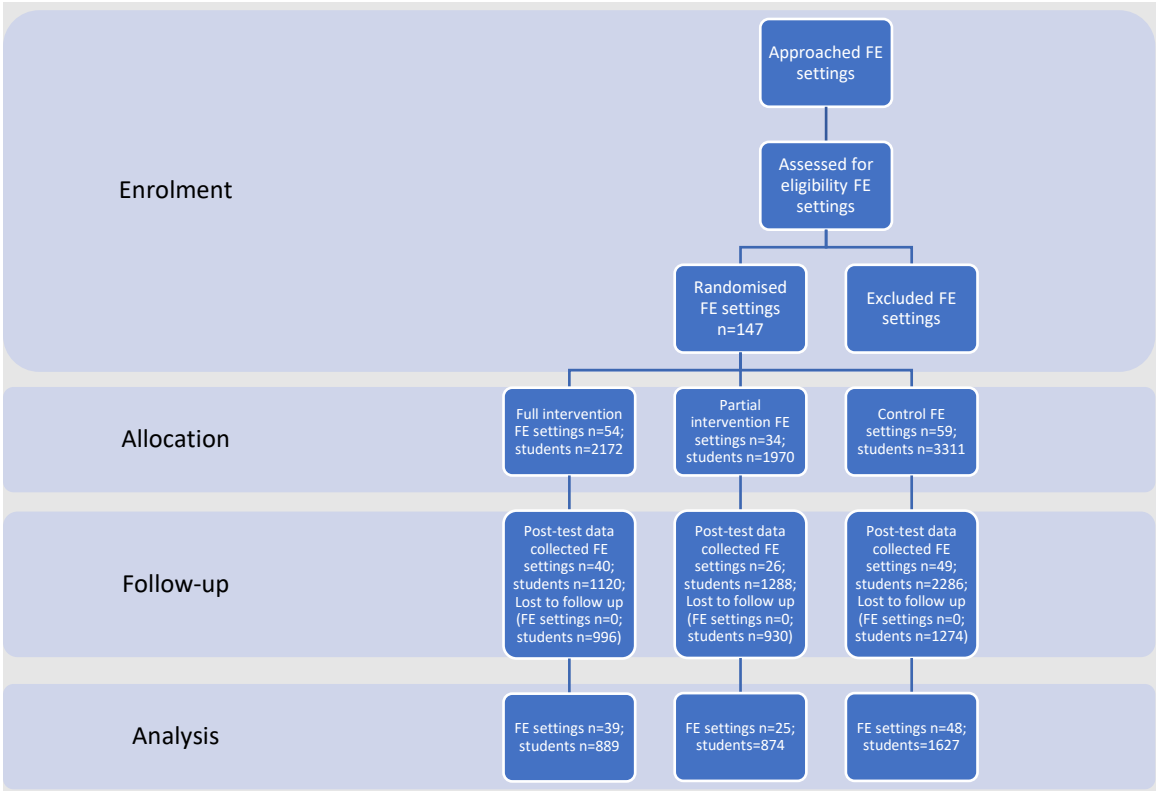
Table 11 Teaching logs

	Lesson 1	Lesson 2	Lesson 3	Lesson 4	Lesson 5
Number of classes (Full intervention)	118	127	238	227	174
Number of classes (Partial intervention)	110	98	107	101	84

Impact evaluation results

The trial was heavily impacted during the second year of the COVID-19 pandemic, and while colleges remained open during the period with no further national or regional lockdowns, student and staff absences were a frequent occurrence, and in some cases, the additional workload became difficult for a number of participants, and as a consequence, settings were withdrawn from the trial. Figure 5 presents the flow of participating students and college settings through each stage of the trial from recruitment, randomisation to follow up testing and analysis.

Figure 5 CONSORT diagram of the trial



Attrition

As demonstrated by Figure 5, student and setting attrition was significant. 7453 students in 147 settings were randomised prior to the start of the trial. By the follow up collection of GCSE post-test data from the FE colleges, 4694 students remained clustered in 115 settings. Due to further missing data within the National Pupil Database extract, this decreased further to 3390 students within 112 settings. As such, the total student attrition was 54.5%, and the total setting attrition was 23.8%.

In terms of differential attrition, the intervention groups were similar in their loss of participants at the student level, with 59% lost to the full intervention group, and 55.6% to the partial intervention group. The control group saw relatively lower attrition at 50.9% although it clearly remained high. The settings level has a greater impact on statistical power given the way in which the trial has been constructed. This again followed a similar pattern of approximately 26-28% of settings being lost to the two intervention groups. As the required participation level was lower, there was a lower level of setting attrition in the control group with 18.6% of settings lost to analysis.

Table 12 Student and setting level attrition for primary outcome (Source: ONS)

		Full Intervention	Partial Intervention	Control	Total
Number of Students	Randomised	2172	1970	3311	7453
	Analysed	889	874	1627	3390
Number of Settings	Randomised	54	34	59	147
	Analysed	39	25	48	112
Student Attrition	Number	1283	1096	1684	4063
	Percentage	59.0%	55.6%	50.9%	54.5%
Setting Attrition	Number	15	9	11	35
	Percentage	27.8%	26.4%	18.6%	23.8%

Minimum detectable effect size

As discussed earlier, the minimum detectable effect size (MDES) was calculated using the R package *PowerUpR* (Bulus et al., 2021) at the protocol, randomisation, and analysis stage and are presented in Table 10. The trial was designed, in ideal conditions, to achieve a MDES of 0.24 on the standard deviation scale between the full intervention group and the control group, and 0.28 between the partial intervention group and the control group. This has been set out earlier in the report and involved the following assumptions – an alpha level of 0.05, a beta level of 0.8, a two-sided hypothesis, an average cluster size of 65 students, an inter-class correlation (also known as Intracluster Correlation Coefficient or ICC) of 0.2 based on previous EEF trial in which the University of Nottingham acted as the developer, and finally a KS2 pre-post correlation of 0.55 (an R^2 of 0.3), with a group-level correlation coefficient of 0.387 (an R^2 0.15).

Due to better-than-expected recruitment, resulting in 54 settings being randomised to the full intervention group, 34 to the partial intervention group, and 59 to the control group. As such, the minimum detectable effect size was updated to 0.23 and 0.26 for the full intervention vs. control, and the partial intervention vs. control contrasts respectively. As with the design stage, the assumptions from the design stage were used.

Given the discussion above regarding attrition, the MDES at the analysis was higher at 0.245 for the full intervention vs. control, and 0.279 for the partial intervention vs. control. We altered the assumptions around ICC, average cluster size and pre-post correlation. For the ICC we observed a correlation of 0.14. For the average cluster size, we used 29 students per cluster for the full intervention vs. control analysis, and 34 per cluster for the partial intervention vs. control analysis. Pre-post-test correlation was also observed to be significantly weaker and so the MDES calculations were now based on a pre-post-test correlation of 0.25 (R^2 of 0.063) and a group-level correlation of 0.23 (R^2 of 0.053). The mitigating factor in keeping the MDES from rising substantially was the lower observed inter-class correlation coefficient of 0.14, rather than the original assumption of 0.2.

Table 13 Minimum detectable effect size at different stages (Source: ONS)

	Design		Randomisation		Analysis	
	Full intervention vs. control	Partial intervention vs. control	Full intervention vs. control	Partial intervention vs. control	Full intervention vs. control	Partial intervention vs. control
Minimum Detectable Effect Size (MDES)	0.239	0.277	0.227[7]	0.259	0.245	0.279
Total Number of settings	100	80	113	93	87	73
Intervention	50	30	54	34	39	25
Control	50	50	59	59	48	48
Total Number of students	6500	5200	5483	5279	2516	2501
Intervention	3250	1950	2172	1970	889	874
Control	3250	3250	3311	3309	1627	1627

Baseline characteristics

Table 14 provides a summary of the FE college for all that were randomised in July 2021, and student baseline characteristics at the analysis stage with the data matched by the DfE's data sharing team. At the setting-level, most FE settings were based in urban settings with only 8 that were in rural locations. In terms of Ofsted rating at the point of randomisation, while there was small variation across the three treatment arms in terms of the proportion of settings rated outstanding and good, all arms three had very similar proportions of settings rated as requiring improvement or inadequate. In terms of the quantitative measures, all three arms had an average of between 94 and 100 students resitting GCSE Mathematics, although the full intervention arm did have much more substantial deviation around the mean. The average number of mathematics groups was between 6 and 7, and on the progress score for Mathematics, all arms had an average measure of -0.04, albeit with some relatively small differences in the standard deviation between the three arms.

At the student-level the equivalence between the two contrasts was generally good. For the full intervention group versus the control group, there was a slight difference in the proportion of males and females with 47% of the sample being male in the full intervention group, and 50% of the sample being male in the control group. However, they were very similarly distributed on the disadvantage measures of FSM6 and FSM-all eligibility, with 36% of students in each group being eligible on the FSM6 measure, and 48% of

[7] The difference in pupil numbers between the partial intervention and the control groups does potentially inflate MDES by a small margin – to within the region of 0.228-0.23.

students with the FSM all measure. For the quantitative variables, there was a small difference on the KS2 scaled mathematics point score of 0.04, and a larger difference, of just under 0.1 standard deviation units on the KS4 Attainment 8 score measure.

For the partial intervention group versus the control group, the proportion of male students was lower with 43% of the sample being male, but like with the primary contrast, they were very similarly distributed on the two disadvantage measures. For both groups, there were 36% of the sample that were eligible on the FSM6 measure, and a minor difference on the FSM all eligible measure of 46% of the partial intervention group sample versus 48% in the control group. For the quantitative variables, there was a small difference on the KS2 scaled mathematics point score of 0.07, and a very small difference of 0.02 standard deviation units on the KS4 Attainment 8 score measure.

Table 14 Setting and Student-level baseline statistics. Settings level baseline characteristics were measured at randomisation. Student-level baseline characteristics were measured at the analytical stage. (Source: ONS)

Further Education Setting Categorical	Full Treatment n (missing)	Count (%)	Partial Treatment n	Count (%)	Control n	Count (%)
Urban vs. Rural	54	1 (2%) Rural 53 (98%) Urban	34	3 (9%) Rural 31 (91%) Urban	59	4 (7%) Rural 55 (93%) Urban
Ofsted	54	5 (9%) Outstanding 36 (67%) Good 11 (20%) Requires Improvement 1 (2%) Inadequate 1 (2%) Missing	34	2 (6%) Outstanding 25 (74%) Good 7 (21%) Requires Improvement	59	3 (5%) Outstanding 43 (73%) Good 13 (22%) Requires Improvement
Further Education Setting Continuous	**Full Treatment n**	**Mean (SD)**	**Partial Treatment n**	**Mean (SD)**	**Control n**	**Mean (SD)**
Average Number of Students Resitting Maths	54	99.6 (90.4)	34	93.8 (50.3)	59	99.2 (69.1) 1 Missing
Average Number of Mathematics Groups	54	6.6 (6.0)	34	5.9 (3.1)	59	6.7 (4.1) 1 Missing

Progress in Mathematics	54	-0.04 (0.18) 3 Missing	34	-0.04 (0.10)	59	-0.04 (0.14) 1 Missing	
Student-level Categorical	Full Treatment n	Count (%)	Partial Treatment n	Count (%)	Control n	Count (%)	
Gender	889	416 (47%) Males 473 (53%) Females	874	379 (43%) Males 495 (57%) Females	1627	814 (50%) Males 791 (49%) Females 22 (1%) Missing	
FSM6	889	323 (36%) FSM 518 (58%) Non-FSM 48 (5%) Missing	874	316 (36%) FSM 526 (60%) Non-FSM 32 (4%) Missing	1627	580 (36%) FSM 981 (60%) Non-FSM 66 (4%) Missing	
FSM All	889	426 (48%) FSM 415 (47%) Non-FSM 48 (5%) Missing	874	400 (46%) FSM 442 (51%) Non-FSM 32 (4%) Missing	1627	778 (48%) FSM 793 (49%) Non-FSM 56 (3%) Missing	
Student-level Continuous	Full Treatment n	Mean (SD)	Partial Treatment n	Mean (SD)	Control n	Mean (SD)	Effect size
KS2 Maths Points Scaled	889	-0.04 (0.98)	874	-0.15 (1.01)	1627	-0.08 (0.98)	0.04 (Primary) 0.07 (Secondary)
KS4 Attainment 8	889 (10 missing)	31.13 (9.02)	874 (10 missing)	30.41 (9.23)	1627 (25 missing)	30.25 (8.76)	0.099 (Primary) 0.02 (Secondary)

Outcomes and analysis

The following section is divided by between the primary and secondary contrasts – the full intervention group versus the control group, and the partial intervention group versus the control group respectively. It will discuss the distributions of the core outcomes and pre-test. It will also present the findings of the primary outcome for the full sample and the FSM subgroup, and for the secondary outcomes of the Mastery subscale and the GCSE Mathematics grade score. Following that, the analysis will present the intra-class correlation, missing data imputation section, and lastly the compliance analysis.

Primary contrast

For the full intervention vs. control contrast, the primary outcome of interest was the average difference on the standardised GCSE raw score. The distributions of post and pre-test scores were presented in Figure 2 and Figure 4.

The primary and FSM interaction models were assessed on the basis of intention-to-treat using a two-level hierarchical model estimated by Bayesian inference. The models were fitted, assessed for convergence and the resulting posterior distributions processed to provide estimated effect size medians, 95% credible intervals, Region of Practical Equivalence (ROPE) and Probability of Direction (pD). These are presented in Table 15.

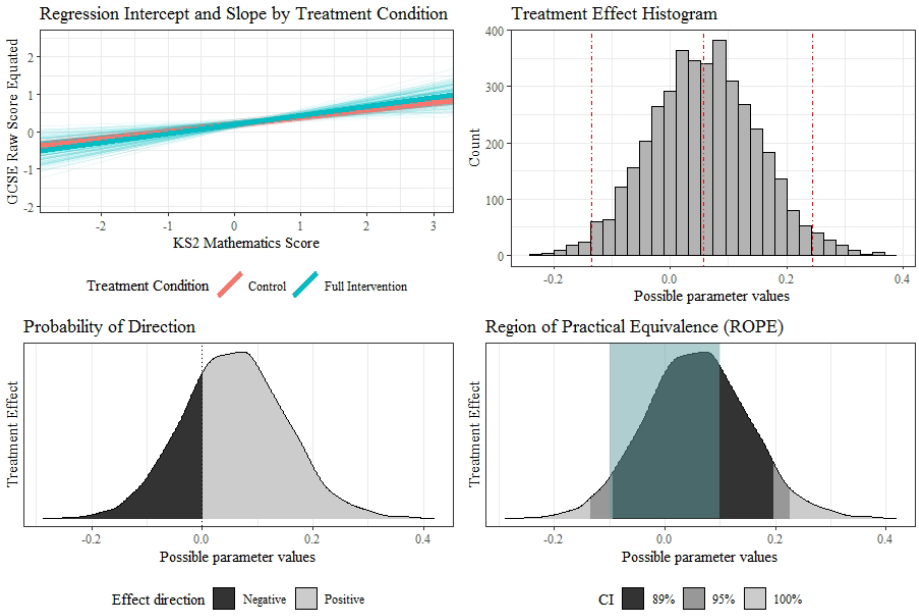
Table 15 Full intervention vs. control primary outcome and FSM interaction model results (Source: ONS)

	Unadjusted Means					
Analysis	Full Intervention	Control	Total n (Treatment/ Control)	Hedges g effect size (95% CI)	ROPE	pD
Full	0.16 (0.80)	0.14 (0.81)	2516 (889, 1627)	0.06 (-0.12, 0.24)	66.1%	0.74
FSM	0.07 (0.83)	0.04 (0.81)	903 (323, 580)	0.11 (-0.10,0.32)	47.0%	0.85

After model fitting and the estimation of the effect size, this resulted in a Hedges *g* effect size difference of 0.06 (-0.12, 0.24). As shown by the pD statistic, 74% of the posterior samples were

positive, indicating weak evidence of a positive effect, albeit not one that is clearly bound away from 0. Interestingly, the ROPE statistic also suggests mixed evidence of a practical effect - only 66.1% of the interval's samples fall between ±0.1 on the effect size scale. The results are further summarised in Figure 6 which presents a regression plot to visualise the small difference between the intervention and control groups. The bold lines represent the median effect, with the faint lines representing the uncertainty in the estimate. The treatment effect histogram presents the range of plausible values on the standard deviation scale and highlights how much the median effect size is shifted away from 0, although not for the entire 95% credible interval. The probability of direction plot highlights the proportion of the effect size distribution which is positive, and lastly, the ROPE plot highlights the mixed evidence with regard to whether the resulting effect is practically different to 0.

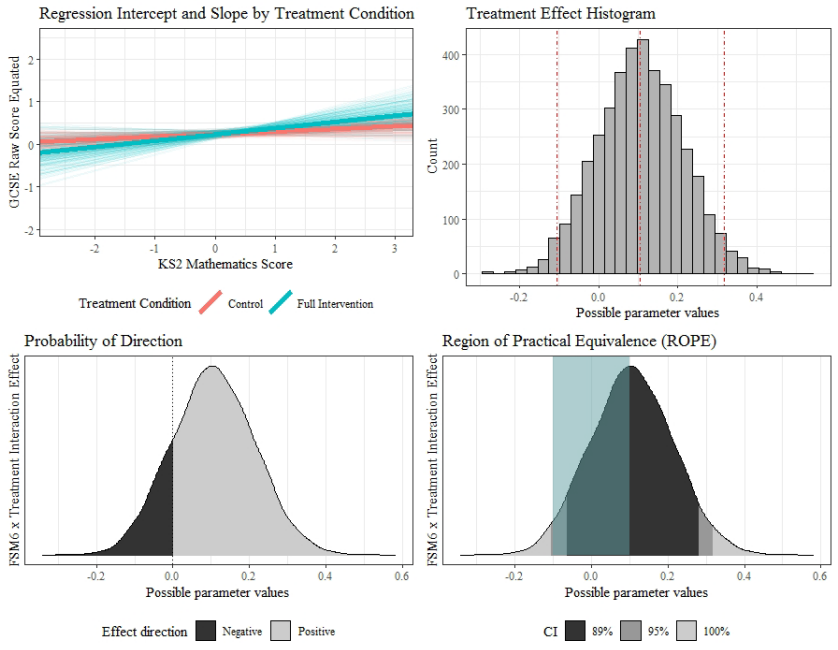
Figure 6 Primary outcome regression plot, treatment effect size histogram, Bayesian probability of direction, and region of practical equivalence on the standard deviation scale, students n=2516 (Source: ONS)



For the FSM interaction analysis, after model fitting and effect size estimation, this resulted in a Hedges *g* effect size of 0.11 (-0.10, 0.32). As before, the 95% credible interval crossed the 0 boundary, but as shown by the pD statistic, 85% of the samples were positive, and only 47% of the interval's samples fell between ±0.1 on the effect size scale. This does suggest mixed evidence of a practical effect, and stronger than the full sample. Given that there still is a substantial uncertainty around these estimates, they should be interpreted very cautiously. In comparison to Figure 6, Figure 7 shows an increased divergence in the two regression lines, although again with some uncertainty. The treatment effect histogram showed the median and 95% intervals shifting further to the right, providing slightly clearer evidence of a potential effect. This was also repeated by the

probability of direction plot and the ROPE plot. The evidence was still mixed regarding being practically significant, but the evidence was clearer than with the full sample.

Figure 7 Primary outcome FSM interaction model regression model, treatment effect size histogram, Bayesian probability of direction and region of practical equivalence on the standard deviation scale, students n=903 (Source: ONS)



As with the primary analysis discussed above, the secondary outcomes were assessed on the basis of intention to treat using a multilevel/hierarchical linear model estimated by Bayesian inference. Recall that Figure 3 depicted the distribution for the item based standardised GCSE Mastery scale.

As reported in Table 16, and in the same manner as the primary outcome, the secondary outcomes all cross 0, indicating no clear evidence of an effect with the existing data collected. However, they all saw positive effect sizes, with the mastery scale for FSM students standing out with a median effect of 0.13 (-0.07, 0.34), having 90% of the posterior distribution being positive, and only 37.8% of the posterior samples within ±0.1 standard deviation units around 0.

Table 16 Full intervention vs. control secondary outcome and FSM interaction model results (Source: ONS)

Analysis	Unadjusted Means		Total n (Treatment/ Control)	Hedges g effect size (95% CI)	ROPE	pD
	Full Intervention	Control				
Mastery Scale	0.11 (0.88)	0.08 (0.88)	2516 (889, 1627)	0.07 (-0.11, 0.24)	64.1%	0.75
Mastery Scale -FSM	0.05 (0.91)	0.00 (0.87)	903 (323, 580)	0.13 (-0.07, 0.34)	37.8%	0.90

Grade Scale	2.24 (1.02)	2.22 (1.02)	2516 (889, 1627)	0.05 (-0.13, 0.23)	69.3%	0.71
Grade Scale - FSM	2.12 (1.04)	2.10 (1.01)	903 (323, 580)	0.08 (-0.13, 0.29)	56.4%	0.77

Secondary contrast

As with the primary contrast analysis, the models were assessed on the basis of intention-to-treat using a two-level hierarchical model estimated by Bayesian inference. The models were fitted, assessed for convergence and the resulting posterior distributions processed to provide estimated effect size medians, 95% credible intervals, Region of Practical Equivalence (ROPE) and Probability of Direction (pD). These are presented in Table 17.

Table 17 Partial intervention vs. control primary outcome and FSM interaction model results (Source: ONS)

Analysis	Unadjusted Means		Total n (Treatment/ Control)	Hedges g effect size (95% CI)	ROPE	pD
	Partial Intervention	Control				
Full	0.22 (0.87)	0.14 (0.81)	2501 (874, 1627)	0.04 (-0.16, 0.25)	67.1%	0.64
FSM	0.11 (0.89)	0.04 (0.81)	896 (316, 580)	0.03 (-0.19, 0.25)	64.4%	0.61

As with the previous analysis, all outcomes crossed 0, with no clear evidence of an effect with the existing data collected as part of the trial. The median effects were smaller at 0.04 and 0.03 for the primary outcome and FSM subgroup analysis, with similar effects shown in Table 18, and the probability of direction did not exceed 0.66 across all analyses, with a value of 0.5 indicating the plausible values of the effect were equally spread between positive and negative.

Table 18 Partial intervention vs. control secondary outcomes and FSM interaction model results (Source: ONS)

Analysis	Unadjusted Means		Total n (Treatment/ Control)	Hedges g effect size (95% CI)	ROPE	pD
	Partial Intervention	Control				
Mastery Scale	0.18 (0.91)	0.08 (0.88)	2498 (871, 1627)	0.03 (-0.17, 0.23)	68.0%	0.64
Mastery Scale –FSM	0.08 (0.94)	0.00 (0.87)	896 (316, 580)	0.04 (-0.18, 0.26)	61.9%	0.62
Grade Scale	2.35 (1.08)	2.22 (1.02)	2498 (871, 1627)	0.04 (-0.15, 0.24)	68.9%	0.66
Grade Scale - FSM	2.20 (1.13)	2.10 (1.01)	896 (316, 580)	0.01 (-0.22, 0.23)	64.4%	0.54

Intra-class correlation

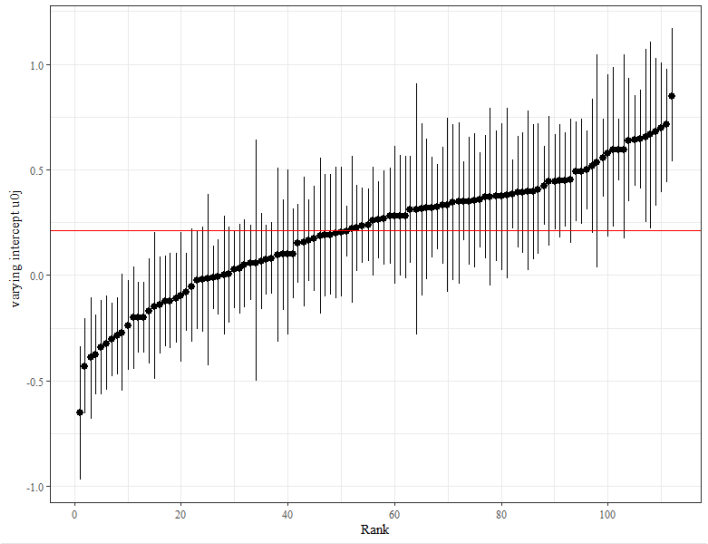
Intra-Class Correlation (ICC) was estimated using an "empty" variance components multilevel model where only the overall intercept and group-level intercepts are estimated. For the primary outcome, it was found that the ICC was smaller than expected indicating less variation at the FE settings level. At the point of randomisation, this was expected to explain approximately 20% of the variation in exam scores. However, at the point of analysis, this was found to only account for 14% of the variation in the exam scores. This result is presented in Table 19 below. The group intercept adjustment predictions - u_{0j} are visually presented in Figure 8 along with their associated credible intervals to provide an indication of uncertainty.

For the GCSE Mastery Scaled Score the ICC was 0.13, and for the GCSE grade score it was 0.14. Finally, for the KS2 Scaled Mathematics Point Score, the ICC was much smaller, only indicating 5% of the variance could be explained by group membership.

Table 19 ICC analysis of primary and secondary outcomes, alongside the KS2 Mathematics pre-intervention attainment measure, student n=3390, (Source: ONS)

Measure	σ^2	$\sigma^2_{FE\ Setting}$	ICC
GCSE Scaled Score	0.59	0.10	0.14
GCSE Mastery Scaled Score	0.70	0.10	0.13
GCSE Grade Score	0.95	0.15	0.14
KS2 Scaled Mathematics Points	0.94	0.04	0.05

Figure 8 Caterpillar plot of Bayesian group-level (FE settings) intercept adjustments on the standard deviation scale, along with 95% credible intervals (Source: ONS)



Imputation analysis

Table 20 below sets out the proportion of fully observed cases, the number of complete cases per variable, as well as notes which highlight any differential missingness across the three groups of the trial.

Table 20 Missing data summary in the trial, students n=7075, (Source: ONS)

Variable	% Fully Observed	Complete Cases per variable	Notes
GCSE Scaled Raw Score	59%	4151	906 Full Intervention, 838 Partial Intervention, 1180 Missing in Control Condition
KS2 Mathematics Points	81%	5728	371 Missing in Full Intervention, 355 Missing in Partial Intervention, 621 in Control Condition
KS4 Attainment 8	91%	6445	177 Missing in Full Intervention, 139 Missing in Partial Intervention, 314 in Control Condition
KS4 Academic Year	89%	6272	245 Missing in Full Intervention, 189 Missing in Partial Intervention, 372 in Control Condition
Gender	99%	7038	37 Missing in Control Condition
FSM6	87%	6189	275 Missing in Full Intervention, 204 Missing in Partial Intervention, 407 Missing in Control Condition

A multilevel modelling of missingness was run to examine whether any covariates might be predictive of missing data on the scaled score primary outcome, as well as the KS2 Mathematics points score. Students scoring one standard deviation higher on the KS2 Mathematics score were 8% more likely to have missing data. This was repeated with EverFSM6 with students from deprived backgrounds being 6% more likely to have missing data. Male learners were 4% less likely to have missing data.

A multiple imputation analysis was carried out using the R package *jomo* (Quartagno & Carpenter, 2022) which applies a joint modelling approach. The imputation model consisted of a two-level set up with students clustered in settings, and included the scaled score, treatment condition, KS2 Mathematics points, EverFSM6, KS4 academic year, and learner gender. An additional auxiliary variable - KS4 Attainment 8 was also included to provide further data on the general ability of the student.

The MCMC algorithm was run for 100,000 iterations for the “burn-in” or warm-up samples, with 4 imputed datasets drawn from the posterior distribution at 10,000 iteration intervals. The resulting parameter distributions were interrogated for convergence using a visual analysis of the MCMC traces, and also via the Gelman and Rubin (1992) Rhat statistic. All parameters displayed full convergence.

The primary outcome and FSM interaction models were then fitted to each of the imputed datasets and the resulting posterior distributions were pooled. The estimated effect size medians, 95% credible intervals, Region of Practical Equivalence and Probability of Direction are presented in Table 21 below.

The results remain consistent with the complete case analysis above. The full intervention vs. control contrast saw a median effect of 0.07 (-0.10, 0.24), with the region of practical equivalence statistic suggesting mixed evidence of an effect, and the probability of direction statistic suggesting that 78% of the posterior draws were in the positive direction. For the partial intervention vs. control contrast, both the median effect and probability of direction suggest that there is less evidence than the primary contrast of a positive difference between the recipient students and the control condition.

For the FSM interaction model there was a marginally stronger effect with a median difference of 0.08 (-0.10, 0.27), 81% of the samples were in the positive direction, and there is mixed evidence of the practical significance of the result. For the partial intervention, there was no evidence of a difference between recipient students and the control condition. The results are presented in Table 21 below.

Table 21 Primary outcome for full sample and FSM interaction based on imputed data, student n=7075 (Source: ONS)

Model	Contrast	Median (θ)	θ-95% Lower CI	θ-95% Upper CI	Region of Practical Equivalence	Probability of Direction
GCSE Scaled Score Primary Outcome Model	Full intervention vs. Control	0.07	-0.11	0.24	65.14%	0.78
GCSE Scaled Score Primary Outcome Model	Partial intervention vs. Control	0.05	-0.15	0.24	68.8%	0.68
GCSE Scaled Score – FSM	Full intervention vs. Control	0.08	-0.10	0.27	58.54%	0.81

Interaction Model						
GCSE Scaled Score – FSM Interaction Model	Partial intervention vs. Control	0.02	-0.19	0.21	68.40%	0.56

Compliance analysis

The compliance measure consisted of the number of lessons delivered as part of the intervention and was measured at the setting level. In the full intervention group, amongst those within the primary outcome analysis, 64% delivered Lesson 1, 69% delivered Lesson 2, 67% delivered Lesson 3, 67% delivered Lesson 4, and finally 56% delivered Lesson 5. This gave a mean compliance score of 3.2 out of 5. In the partial intervention group, 68% delivered Lesson 1, 68% delivered Lesson 2, 68% delivered Lesson 3, 64% delivered Lesson 4, and 44% delivered Lesson 5. This gave a mean compliance score of 3.1 out of 5.

With this data, an instrumental variables model was run and estimated classically due to software limitations within the ONS Secure Research Service environment. The results remained consistent with previous analyses, although with wider confidence intervals. The full intervention saw an estimated effect size of 0.05, and the partial intervention saw an estimated effect size of 0.04.

Implementation and process evaluation results

Introduction

This was a complex programme, involving two main sets of activities: those with the Lead Teachers and those with the Trial Teachers. Each set of activities had some expectations in terms of outcomes, mediated by a set of mechanisms. Recall that Figure 1 summarised the logic model for the programme. This section takes the activities in turn and evaluates them in terms of the intended short-term outcomes and mediating mechanisms.

Table 22 Timeline for the programme and evaluation instruments

Date	Activity	Evaluation instruments
24th March 2021 Online	**Lead Teacher PD1** Lead teachers participate in a professional development day, led by the University of Nottingham.	Online survey Interviews
23rd June 2021 Online	**Lead Teacher PD2** Lead teachers participate in a professional development day, led by the University of Nottingham.	Online survey Interviews
6th/8th and 13th/15th October 2021 *Mornings*, Online	**Mastery 1** Lead teachers and trial teachers (both full intervention and partial intervention groups) participate in professional development on teaching for mastery, led by the University of Nottingham.	Online survey (anonymous) Interviews (case study teachers and Lead Teachers)
6th and 8th October 2021 *Afternoons*, Online	**Lesson study** Lead teachers and trial teachers (full intervention group) participate in professional development on lesson study and cluster meetings (two sessions available with each participant to attend one), led by the University of Nottingham.	Online survey (anonymous) Interviews (case study teachers and Lead Teachers)
15th October 2021 *Afternoon*, Online	**Cluster Meeting 0** Lead teachers and (full intervention group) participate in Cluster Meeting 0, to prepare for teaching Lesson 1 and Cluster Meeting 1, led by lead teachers.	Online survey (anonymous) Interviews (case study teachers and Lead Teachers)

Autumn term 2021	**Lessons 1 and 2** Trial teachers (both groups) teach lessons 1 and 2 in Windows 1 and 2. Window 1: 1st to 26th November 2021 Window 2: 29th November to 17th December 2021	Post-intervention survey Lesson observations (case study teachers)
Autumn term 2021	**Cluster Meetings 1 and 2** Teachers in the full-intervention group participate in Cluster Meetings 1 and 2 (full intervention group), led by lead teachers.	Cluster meeting observations (case studies) Interviews (case study teachers and Lead Teachers)
5th January 2022	**LT PD3** Lead teachers participate in a professional development day, led by the University of Nottingham.	Online survey (anonymous) Interviews (case study Lead Teachers)
12th and 14th January 2022	**Mastery 2** Lead teachers and trial teachers (both groups) participate in professional development on teaching for mastery, (two sessions available with each participant to attend one), led by the University of Nottingham.	Online survey (anonymous) Interviews (case study teachers and Lead Teachers)
Spring term 2022	**Lessons 3, 4 and 5** Trial teachers (both groups) teach Lessons 3, 4, and 5 in Windows 3, 4 and 5. Window 3: 6th to 28th January 2022 Window 4: 31st January to 4th March 2022 Window 5: 7th March to 1st April 2022.	Post-intervention survey Lesson observations (case study teachers)
Spring term 2022	**Cluster Meetings 3, 4 and 5** Teachers in the full intervention group participate in Cluster Meetings 3 (online), 4 (online) and 5, led by lead teachers.	Cluster meeting observations (case studies) Interviews (case study teachers and Lead Teachers)

Findings

All professional development took place in line with the intended models as evidenced by professional development session logs and the timeline for the professional development sessions found in Section 1.1. Tables 23–25 provide details of Lead Teacher and Trial Teacher participation in these sessions.

Table 23 Attendance of Lead Teacher sessions

	Full attendance	Partial attendance	Non-attendance
LT PD 1&2	8	1	1
LT PD 3	7	0	3
Lesson study PD	10	0	0

Table 24 Attendance of Trial Teacher sessions

	Full attendance	Partial attendance[8]	Non-attendance
PD Mastery 1&2	52	8	8
PD Mastery 3	63	3	2

Table 25 Attendance of cluster meetings

	Attendance	Non-attendance
Cluster 1	34	10
Cluster 2	36	8
Cluster 3	38	6
Cluster 4	32	12
Cluster 5	30	14

Lead Teachers (short term outcomes)

Lead Teachers engaged very well with all three PD sessions (all online). Their PD evaluations (survey data) showed that the majority were very engaged with the presentations and activities (Table 26). No LTs reported that they were mildly engaged or not engaged.

Table 26 The extent to which Lead Teachers engaged in the PD activities

	Very engaged	Moderately engaged	Total
PD 1	8	2	10
PD 2	8	2	10
PD 3	6	1	7

[8] Trial Teachers who could not attend any of the PD days were invited to a twilight catch-up session

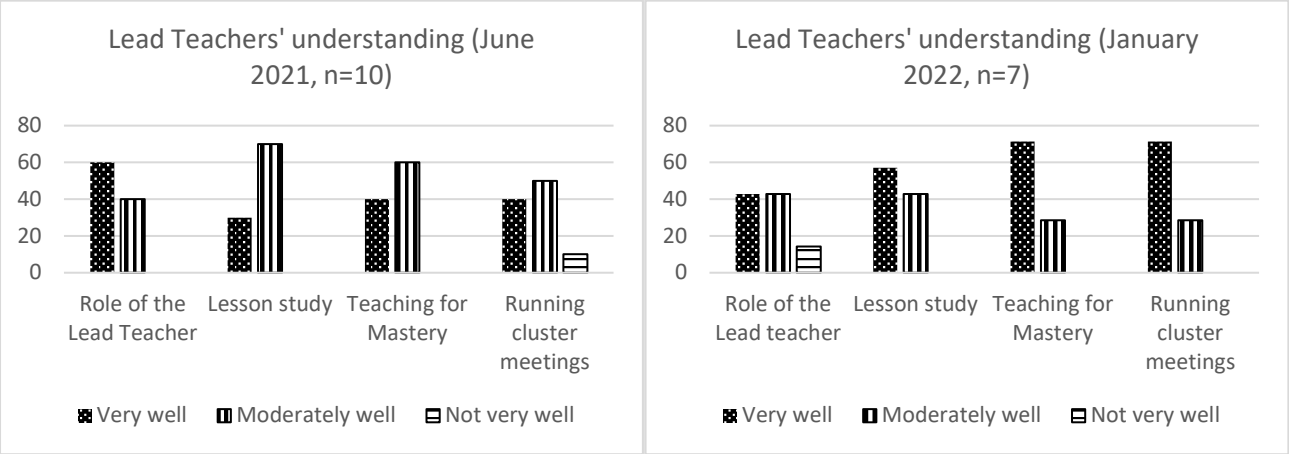
Where engagement was moderate, two Lead Teachers explained that they had technical difficulties, one stating, for example that “I wasn't able to properly engage because my internet connection was very bad, and there was a huge lag - couldn't follow the discussions” (LT4, PD3). One explained he was ill and found it difficult to concentrate, another that he felt under-prepared and the third that he found himself “starting to switch off slightly during the talk on dialogic teaching” (LT1, PD1).

Evidence of a high level of engagement was supported by comments in the PD evaluation surveys such as “I participated in all activities and was very active in discussions” (LT 7, PD Day 2) and “short sharp activities support engagement, group or pair activities support engagement” (LT 3, PD Day 3)).

These were online sessions (due to Covid) but Lead Teachers stated that they were engaged because the sessions were well structured and that the balance of activities worked well. There was evidence that Lead Teachers engaged because they were well motivated “I was keen to learn as much as possible from the activities so was happy to engage and contribute”(LT1, PD Day 1).

For the short- and medium-term outcomes, Lead Teacher learning concerns knowledge of the lesson study process; understanding of the Lead Teacher role; knowledge and understanding of teaching for mastery in FE. There is evidence from the Lead Teacher professional development evaluation surveys and case study Lead Teacher interviews that these outcomes were met. For example, in the evaluation following the second and third professional development sessions, the teachers reported understanding these and their role generally very well or moderately well.

Figure 9 Lead Teachers' TfM understanding in June 2021 and in January 2022



The reality is perhaps more nuanced than the data here suggests. For example, one Lead Teacher explained: “I have answered about my understanding of; the role of a lead teacher, cluster meetings, teaching for mastery as being moderately good as I feel that I am learning more and more. I hope that I will continue to learn.” (LT2, professional development evaluation survey, LTPD

3). This Lead Teacher was a case study Lead Teacher, and observations of the cluster meetings he ran and the two TfM lessons he taught at cluster meetings demonstrate good understanding of the TfM approach and running cluster meetings.

Lead Teachers (mediating mechanisms)

The mediating mechanisms related to the Lead Teacher activities and learning concern running cluster meetings and facilitating the learning of FI teachers. Lead Teachers were provided with detailed guidance about running the cluster meetings, which included checklists, support for the lesson study observations and discussions, and an introduction to the next lesson to be taught (see Appendix A for further details).

Observations of cluster meetings (case studies A, B, C) suggest that the Lead Teachers followed the guidance closely, deviating only on the order of events and the timings of discussions, in order to fit in with the context of the host teacher and college.

Generally, the Lead Teachers implemented the cluster meetings in line with the model in the resources. This researcher observation of one cluster meeting provides some evidence:

> *The LT followed the guidance closely. This provides a structure for the discussion, carefully going through the different phases of the lesson and asking questions designed to deepen the teachers' understanding of the lesson design. The teachers engaged with the discussion, offering their views on the design and how they thought their students would respond. (LT3_CM2).*

The Lead Teachers reported that the Lead Teacher guidance was very helpful, explaining that reading the guidance and then using the PowerPoint in the cluster meetings meant that they remembered to cover everything and to ask the questions that probed the teachers' understanding of the lesson design. One Lead Teacher explained:

> *Having the lead teacher handbook,.., I I love those, I can't tell you. Whoever writes them, I wanna hug them. Because sometimes if I didn't have that, I think I probably I would like to think there be a couple of things I'd miss. You know, I'm quite honest so they're really useful. (LT2_Int2)*

Trial Teachers' learning (mediating mechanisms)

All trial teachers were required to take part in two Teaching for Mastery PD sessions (see Section 1.1 for the scheduling of these, Table 24 for attendance and Appendix A for details of the content). In addition, TTs in the full-intervention group took part in professional development related to lesson study and cluster meetings. This is discussed separately below.

Trial Teacher professional development

The mastery PD for all full- and partial-intervention teachers aimed to develop teachers' understanding of the Teaching for Mastery Key Principles for FE; how the sample Teaching for Mastery lessons were designed and how they should be used; and how they might adapt their own teaching so that it becomes more aligned with Teaching for Mastery approaches. Appendix B provides a description of this professional development.

The trial teachers were asked to complete professional development evaluation survey after the first event (Mastery 1, October 2021, 38 responses) and after the second (Mastery 2, January 2021, 57 responses). All trial teachers were asked to complete a post-intervention survey in April 2022 (58 responses). Case study teachers were also interviewed about their experience of the professional development.

Trial teachers from both the full- and partial-intervention groups engaged well with the Teaching for Mastery PD sessions they attended with 31/34 (91%) reporting that they were very or moderately engaged (Table 27).

Table 27 The extent to which Trial Teachers engaged in the PD activities

	Very engaged	Moderately engaged	Mildly engaged	Not at all engaged
Mastery 1	61.76%	29.41%	8.82%	0%
Mastery 2	71.15%	23.08%	3.85%	1.92%

Their comments indicated that the interactive activities, such those using online collaborative activities on Desmos[9] and Jamboard[10,] were particularly engaging and that the small group discussions were valuable to collaborate and exchange of views with other teachers about the lessons "It was great to hear other teachers' comments on these" (PD evaluation, anonymous). Interviews with the case study teachers provide more detail, with one, for example, explaining

> *"[The most helpful thing] was going through the exercises. Yeah, so you know we went through the worksheet, didn't we? And then we went through the exercise in Desmos. Actually doing it and then questioning the sort of processes as you do it … [In the breakout room] It was good to get the people that was there and get their opinions as well and then*

[9] Desmos teacher is a platform through which interactive online activities can be provided. The teachers were assigned to small online groups and worked collaboratively through the activities. An example activity can be found here.
[10] Jamboard is a cloud-based app that enables visual collaboration between users in real-time. An example activity can be found here.

have a professional discussion on what the worksheet was supposed to achieve. For me, that was great." (TT14_Int1).

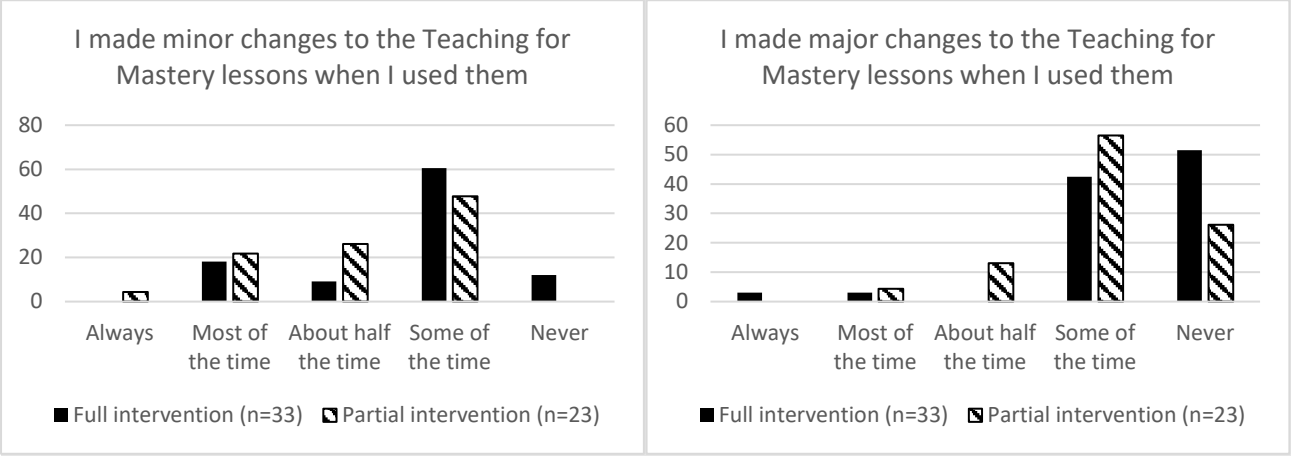
Lessons

A second strand of professional development for all Trial Teachers involved teaching at least five of the seven lessons they were given. The section on dosage, below, provides details about how many lessons were taught and Appendix A provides a description of the resources provided for the teachers.

The Mastery 1 and Mastery 2 professional development emphasised to teachers that they should follow the lesson plans closely, but included discussion about the sorts of adaptations they might make to the lessons either in their planning or in the moment in the classroom.

In the post intervention survey (April 2022), teachers were asked about changes they made to the lessons. This provides information about fidelity of implementation. Figure 10, below, indicates that the majority of teachers made small changes, but most of them (78%) did not do so all or even most of the time. There were small differences between the full intervention and partial intervention groups.

Figure 10 To what extent Trial Teachers made changes to the TfM lessons



When they were asked how often they made major changes, only 5% suggested they did this always or most of the time, with 95% stating that they made changes about half the time, some of the time or never. Figure 11 shows the differences between the two groups, indicating that twice as many full intervention teachers (52%) as partial intervention teachers (26%) never made major changes.

Observations of the case study teachers teaching the lessons provide a mixed picture. Of the eight case study teachers (six full intervention and two partial intervention), one (full intervention) made

major changes to the lesson plans, five made minor changes and two made very minor changes or none at all.

The one who made major changes made it clear that he did not like the lessons, or the Teaching for Mastery approach. He taught the lessons, broadly in-line with the plans, and made multiple small changes and some bigger ones, as described below in the case-study write-up:

> *In most lessons, the teacher only broadly followed the lesson plan. In Lesson 2, for example, he followed the plan but: he took about half the time intended, intervened in student pair activity and discussed specific cards with the whole class, said the first time he taught the lesson, he had given out only the fraction cards to start with and that had worked well, but the second time he had given out the diagram cards as well. He did not explain why. He finished almost all lessons early, skipping either some activities or discussion.*
>
> *He usually made some changes, for example creating extra handouts. In Lesson 1, for example, he said that his students would find it too difficult to work on mini wbs on their desks with the information they needed on the board/screen so he had created a handout for them with all the questions they had to answer. (TT60)*

Of those that made minor changes, observation and interview data suggest that most teachers believed that they followed the lesson plan closely but that almost all made some small unplanned changes particularly in the timings, sometimes spending longer on the introduction to the lesson and usually allowing less time for the student pair-work than indicated on the plan. This is described, for example, in the write up of one full intervention teacher case studies:

> *The teacher adhered quite closely to the lesson plans but made some adjustments to timings, such as shortening the time for paired discussions… Often the time allocated for paired discussion was shorter than in the lesson plan and the teacher sometimes intervened with explanations before the students had sufficient time for productive struggle. (TT114, Case study write-up)*

Some made deliberate changes before the lesson, but these were usually minor and did not change the sense of the lesson. One teacher, for example, provided concrete resources (sheets of paper) to help students visualise what carpet on a roll looked like (IPE_lessonobsform_lesson3_TT67_R1_3101220).

Observations also suggested that while teachers believed that they were following a Teaching for Mastery approach, they sometimes found it difficult to allow discussions in the whole class and

between students to develop fully, as intended. This was observed in lessons of both the full and partial intervention groups, as described below.

> *Although she believed she had adapted and embraced the Teaching for Mastery approach, there was evidence that she found it difficult to move away from a teacher-focussed style and use paired work and class discussions as intended. Opportunities to develop student understanding through exploration and discussion of the tasks were often missed and it was difficult to build on prior knowledge because this was also not explored sufficiently. She did however develop understanding of how to teach with a more connected approach and planned to continue developing this rather than teaching isolated topics. (TT33, PI group)*

> *She generally followed the lesson plan, but intervened with students more, and more often, than suggested by the lesson plan. She appeared to be keen to help the students, usually intervening before they had time to 'struggle'. It sometimes seemed that she provided explanations even though the students did not request them. (TT67, FI group).*

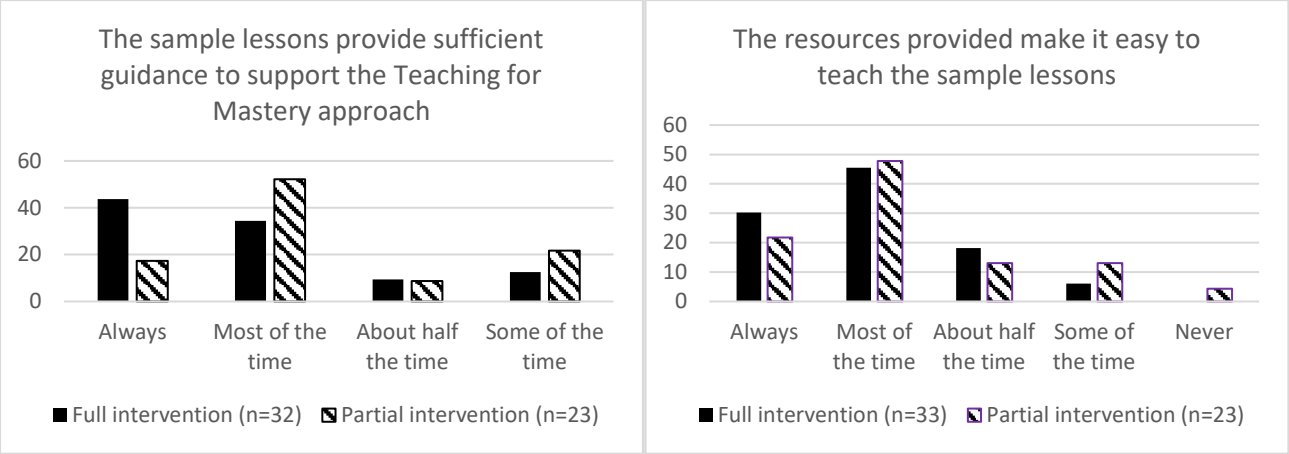
On the other hand, there were two case study teachers, both in the full intervention group, who did appear to embrace a Teaching for Mastery approach fully. They prepared very thoroughly for the lessons and appeared to have an excellent understanding of how the lesson was designed and what it aimed to achieve. They both followed the lesson plan very closely and demonstrated teaching in-line with the five key principles, particularly valuing student contributions, encouraging whole class discussion and allowing enough time for students to work on their own, as described in the excerpt from the write up of the case study teacher TT149:

> *The teacher adhered closely to the timings suggested on the lesson plans and gave ample time for both paired work and class discussion, which allowed the students to explore the mathematical structures. … She refrained from giving answers during paired work but asked for explanations of their thinking and questioned them about their responses to probe further into their thinking… Students were encouraged to try different approaches in the lessons observed and to offer alternative ideas in class discussions. The teacher allowed them space to explore these and asked what they had learned from specific methods that were new to them. The teacher frequently checked on understanding by questioning students about their thinking and probing further into their responses. (TT149)*

The lesson plans and other resources given to the teachers were designed to support teachers in using the lessons. Generally, they reported that the lessons provided enough guidance to support them in using a Teaching for Mastery approach, as shown by Figure 11. Many more full intervention teachers than partial intervention teachers selecting 'always', while more partial intervention

teachers chose ‘most of the time’. This may suggest that the lesson study introduction to the lessons, for the full intervention teachers, influenced their perceptions of the level of guidance provided. As Figure 11 below demonstrates, 75% of all the respondents selected one of these two options and of the 55 teachers who responded, 5 (just under 10%) selected ‘about half the time’ and 7 (about 13%) selected some of the time. None selected ‘never’.

Figure 11 To what extent the lessons provided sufficient guidance



There was very little difference between the two groups when it came to their views on the extent to which the resources made it easy to teach the lessons. Altogether 73% agreed that the resources made it easy to teach the lessons all or most of the time, and 25% agreed that they made it easy half the time or some of the time (Figure 11). One person, who was in the partial intervention group, said the resources never made it easy. Interestingly, this person generally did not appear to like the approach exemplified in the resources, saying that he or she made major changes to the lessons ‘most of the time’, that students never found the lessons engaging and that he or she would not use the lessons with GCSE classes in the future.

In terms of the use of the lessons, there was very little difference between the two groups when it came to the number of the lessons delivered. Both full intervention and partial intervention teachers delivered 3 out of 5 lessons, on average, and fewer teachers delivered Lesson 5 compared to other lessons (see the impact evaluation results section). However, we argue that the number of lessons delivered should be higher, given there were lessons being taught but not being recorded in the teaching logs. For example, we observed that a teacher did not record their teaching on Lesson 4 and 5. There was also no observable relationship between the number of the lessons delivered and the number of PD or cluster meetings that teachers attended. We observed that some teachers used the lessons in one class but not in another.

Further Trial Teacher professional development – Full intervention group only

Full intervention teachers took part in the 'Lesson study' professional development which aimed to prepare them for taking part in cluster meetings.

Of the 29 full intervention teachers who completed an anonymous evaluation of this session, 28 reported that they understood their role very well (11) or moderately well (17). Commonly, they commented that they had a good idea of what was involved but that they had some questions or needed to clarify some details. Some suggested that they would be clearer once the intervention was running. For example, one teacher commented:

> *After the training today and reading the Trial teacher handbook, I feel I have a good understanding of what is required of me. Once it all starts fully I will have a much better understanding of the whole process. (Anonymous)*

In a second afternoon session (online, October 15th), the Lead Teachers met with their cluster groups for the first time. In this session they worked through the materials designed to introduce the teachers to Lesson 1. All teachers had the opportunity to work in small collaborative groups through the main student task for this lesson using a Jamboard version. Teachers particularly appeared to value meeting others in their cluster and working through the task. Their comments included:

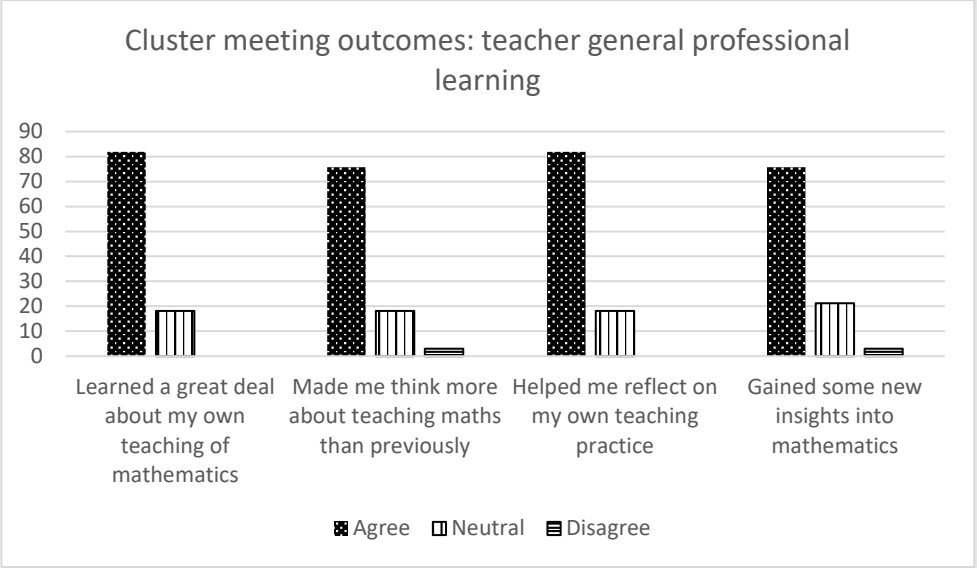
> *All the sessions have been useful, but meeting other teachers and discussing our experiences so far. (Anonymous)*

> *Activities in breakout room gave me a better understanding of the activities that I am expecting in the lesson. But need a bit more help understanding what we are to do. (Anonymous)*

The full intervention teachers participated in five cluster meetings, at which they observed one member of the cluster teaching the lesson. After the lesson, the whole group discussed the lesson and prepared for teaching the next lesson. A description of what happened at cluster meetings can be found in Appendix B.

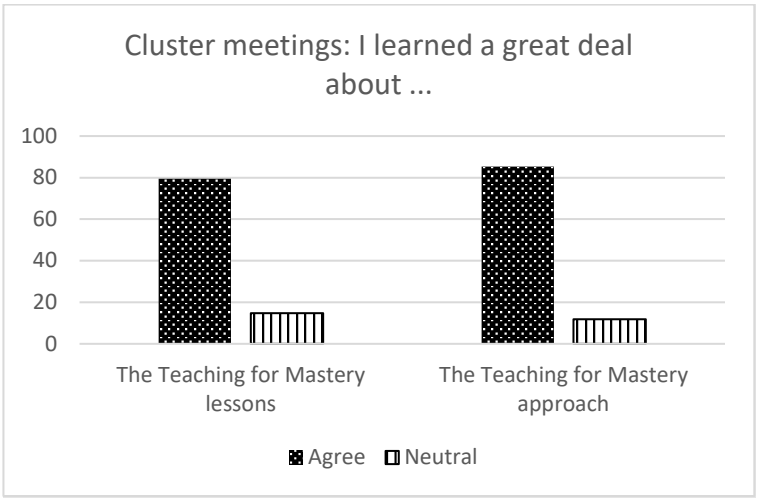
In terms of the Trial Teachers' experiences of the cluster meetings, generally the feedback was very positive. Of the 34 teachers who completed the post-intervention survey, 30 agreed that taking part in the cluster meetings was a valuable experience. They were asked about what they learned at cluster meetings. A set of questions related generally to their professional learning and for all questions the majority agreed that learning had taken place, as shown in Figure 12 below.

Figure 12 Trial Teachers' experiences of the cluster meetings



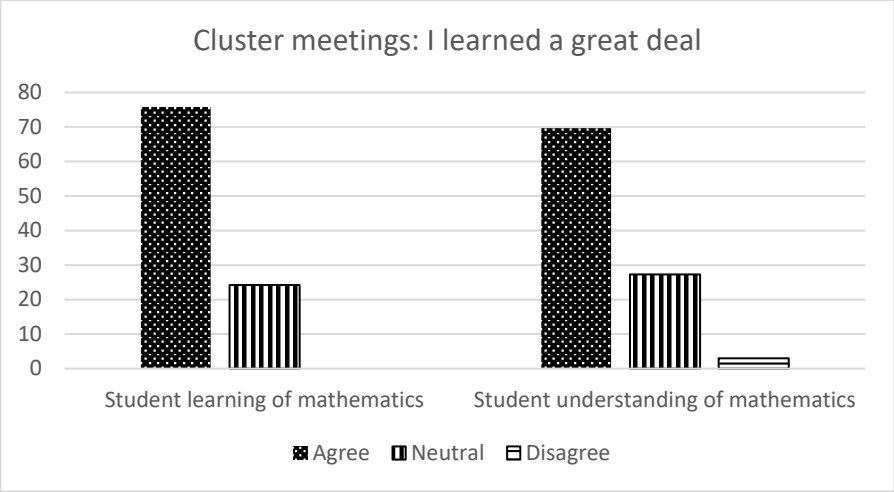
Teachers were asked specifically about their learning with respect to Teaching for Mastery, and for both questions, thereabout 80% of the teachers agreed that they learned about the lessons and the Mastery approach (Figure 13).

Figure 13 Trial Teachers' learning with respect to Teaching for Mastery



In terms of student learning and understanding, most agreed that they had learned a great deal, as illustrated in Figure 14 below.

Figure 14 Trial Teachers' learning with respect to student learning and understanding



The teachers appeared to value many aspects of the meetings. They discussed the value of observing someone else's lesson. One, for example, explained:

> *But it is the most amazing experience to be able to watch someone else teach, and I think Andrew's lesson in particular really showed me value of perseverance, particularly with students who appear to be disengaged. So he had a couple of lads who looked like they really didn't, weren't interested and he persevered one to one and monitored the task and persevered. And then you saw this lad, the light bulb go on and all of a sudden he sat up straighter and started writing on his whiteboard. And I looked and I thought, yeah. I mean how many times have I given a prompt and walked away too quickly and I noticed in his teaching just how much time he spent waiting after the prompt and I that was really instructive for me. I thought yeah, I need to take something out of this with 16 to 18s about waiting time and just the way he handled the ones that looked disengaged was really, really helpful for me. So that lesson was amazing. (Case study TT157, Int 2)*

A number of teachers reported on the benefit of observing and discussing different approaches to teaching and learning. For example, when asked what they got out of the post-lesson discussion, one teacher described the value of hearing about how the students other teachers' classes responded to a lesson. Another mentioned that, in the lesson observation, teachers in the cluster group noticed different things and that had made her think about what might happen in her classroom and how she would manage it.

Trial Teacher 149 summed up the value of cluster meetings in her fourth interview:

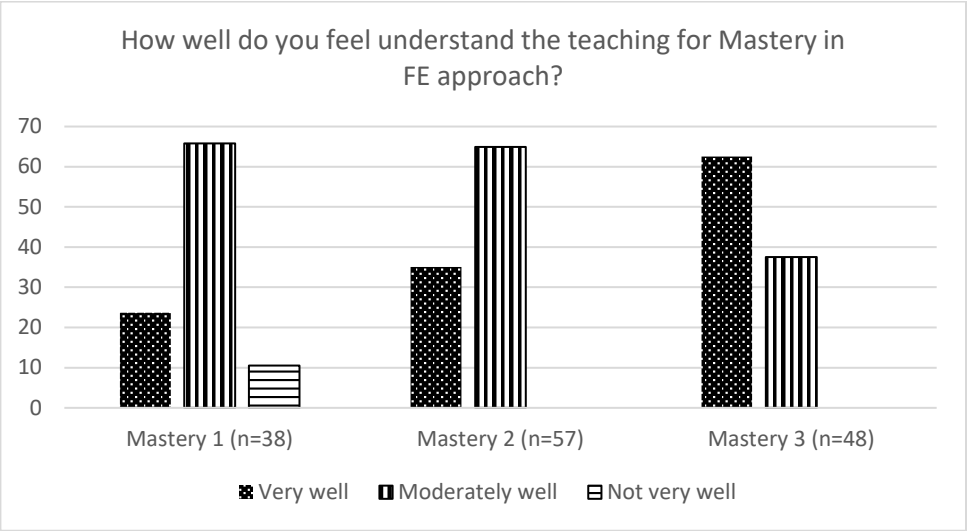
> *It's really about the research and about … let's say the quality of the lesson and the lesson in that way, does it draw out of the student what you what we intended it to do? So the cluster meetings are an absolute fundamental part of the whole.*

Throughout this research project, and I do feel without them, it wouldn't have been as exposed, as beneficial, as successful, as what it has been. They are, you know, they they're really important. (Case study TT149, Int 4).

Mastery

Teachers reported on their understanding of the teaching for Mastery in FE. After Mastery 1, they had not yet taught any of the lessons, but over 20% reported that they had a very good understanding and over 60% reported that they had a moderately good understanding. In later surveys, having taught the lessons, increasing numbers reported having a very good understanding. Figure 15 illustrates this trend.

Figure 15 Trial Teachers' understanding of Mastery over time (in percentage)



The survey included questions about two specific key principles. Teachers were asked how well they understood the ways in which student tasks and representations are used to develop an understanding of mathematical structure; 95% and 100% of those who responded selected very or moderately well after the October and January professional development respectively. The responses to a question about understanding of the ways in which teachers value and build on the prior learning of students were similar (95% and 96% respectively).

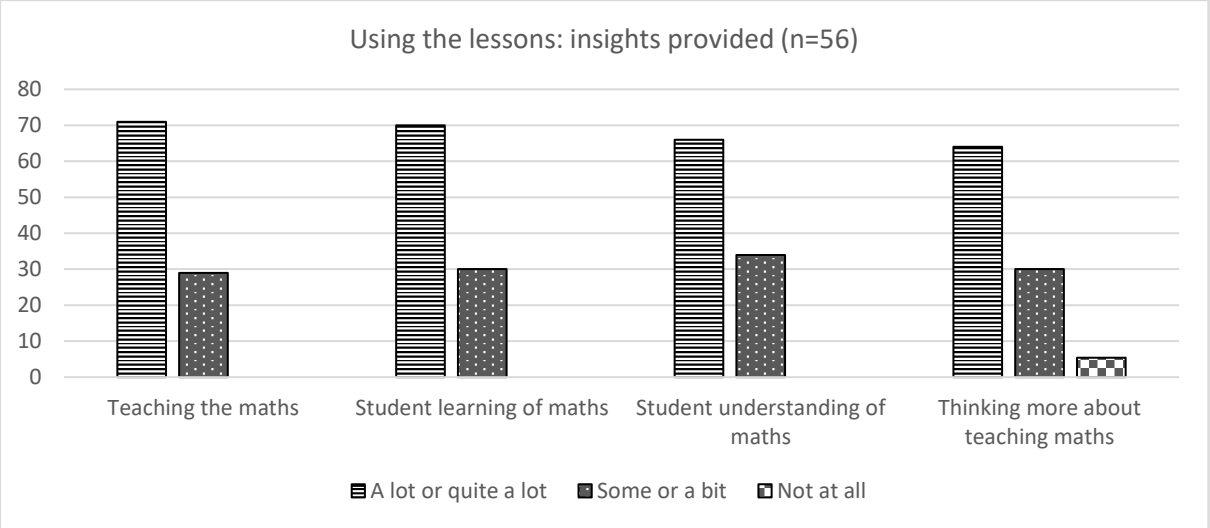
The case study teachers, who were first interviewed after the first two half-days of mastery professional development confirmed that their understanding of teaching for mastery and the mastery key principles had developed over the course of the professional development. One, for example, explained:

From the sessions I then developed an understanding that mastery is about the key principles, it's the principles, and I think that was for something I didn't understand before coming in. I just always assumed it was making things pretty and putting bar models and

using visual representation. So I sort of had an understanding, but I think it was a skewed understanding. (TT114, interview 1, 19/10/21)

The teachers were asked to respond to a series of questions about the extent to which their use of the Teaching for Mastery lessons had provided insights into teaching mathematics, students' learning of mathematics and students' understanding of mathematics; and also the extent to which the use of the lessons had made them think more about teaching mathematics (with 'A lot' at one end of the scale and 'Not at all) at the other). As Figure 16 indicates, the majority of teachers reported that using the lessons had developed their insights and made them think more about teaching mathematics a lot, or quite a lot. Three teachers reported that using the lessons had not made them think more about teaching mathematics.

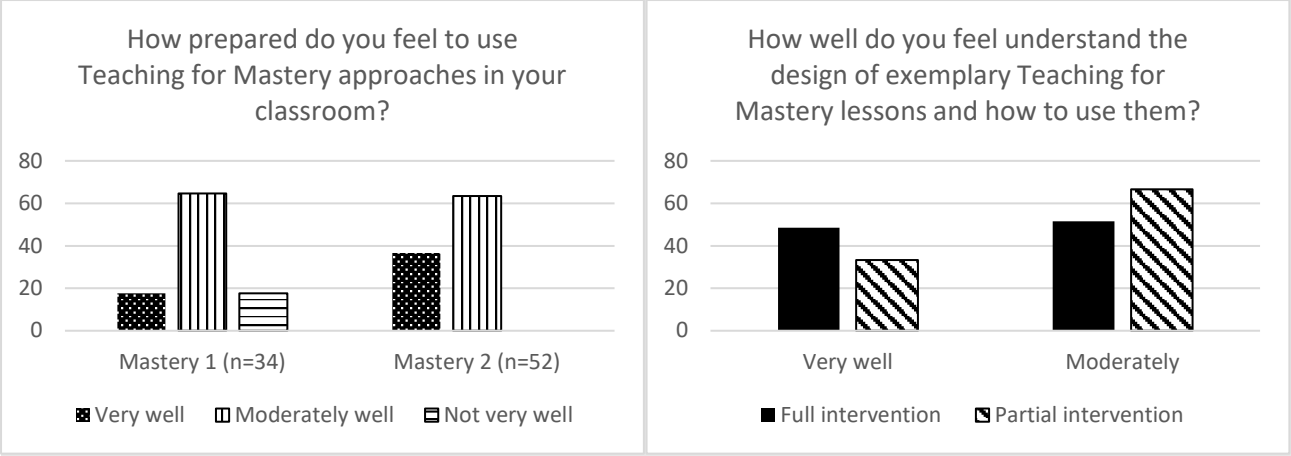
Figure 16 To what extent Trial Teachers' use of the TfM lessons provided insights into teaching



Learning about putting mastery into practice

Teachers were asked about the extent to which they felt prepared to teach the mastery lessons after the first two professional development sessions, and, as Figure 17 demonstrates, almost all felt very or moderately well prepared after the first session (Mastery 1) and by the end of the second session, all felt prepared and a greater number felt very prepared.

Figure 17 Trial Teachers' feelings about TfM



The case study teachers confirmed that they had improved understanding of how to put the Teaching for Mastery principles into practice. One of the teachers, TT14, explained that before the first Mastery professional development session, he had not been confident about Mastery at all, although he had read about it. He said that after the session, he felt that he knew much more, going on to explain what he had learned:

> *It's the approach … of giving the students the time to learn and time to, well, investigate, isn't it really, on what they know already know and then link everything back together. That that was quite important to me. Give them the time to learn or give them the time to investigate and then they will come up with their answers. (TT14_Int1)*

In terms of developing an understanding the design of the lessons, at the end of the intervention both the full and partial-intervention teachers reported that they felt they understood the design of the lessons very well or moderately well.

Figure 18 also shows that just under 50% of the full-intervention teachers reported understanding the design of the lessons very well and just over 50% reported understanding moderately well. For the partial-intervention teachers, however, a smaller number (just over 30%) reported understanding very well, and well over 60% reported understanding moderately well. This suggests that the extra professional development for the full-intervention teachers may have developed their understanding of the design of the lessons.

The professional development also aimed to improve the Trial Teachers' understanding of how to adapt their own teaching so that it became more aligned with Teaching for Mastery. By the end of the intervention, the majority of teachers in both groups reported that they understood very well, or moderately well, how to do this. There were some differences between the full intervention and the partial intervention groups: first, a small number of teachers (2) in the full intervention group,

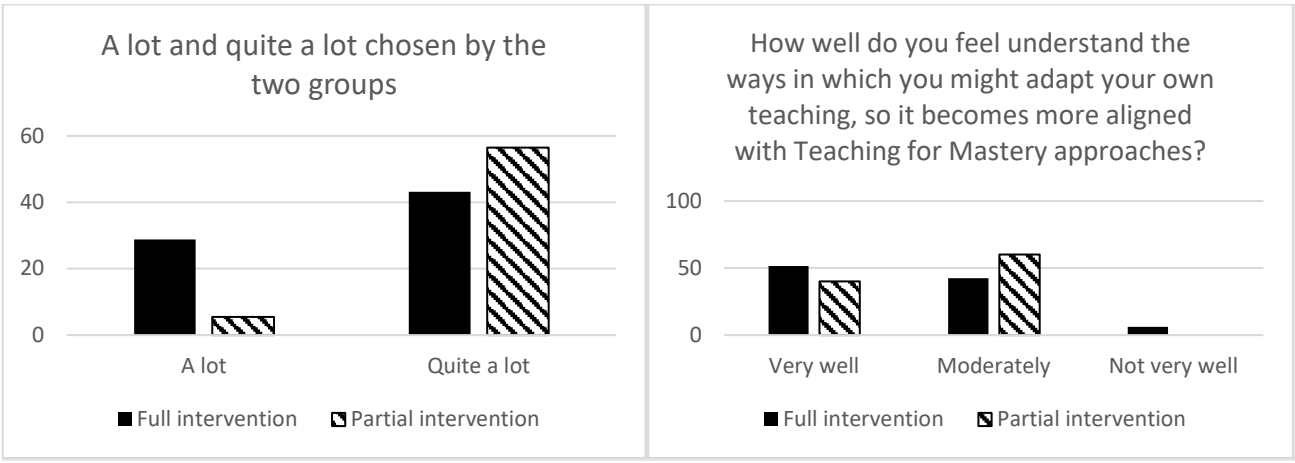
reported that they did not understand very well and second, in the full intervention group more teachers reported a very good understanding than a moderate understanding, whereas in the partial intervention group it was the other way around (Figure 18). This could be explained by the detailed introduction to the lessons the teachers in the full intervention group had in the cluster meetings.

Learning by teaching the lessons

Teachers were asked a set of questions related to the contribution of the use of the lessons to their own learning; for each statement they were asked to select a response, from 'a lot' to 'not at all'. The results (percentages of responses, n=33 for full intervention and n=23 for partial intervention) were similar in the two groups, with the majority selecting a lot or quite a lot, and a much smaller number selecting some or a bit. Two teachers (one full intervention and one partial intervention) selected never. Overall, this suggests that teaching the lessons provides the teachers with valuable learning experiences.

However, the two groups were quite different in terms of their selection of *quite a lot* and *a lot*. If the numbers of *a lot* and *quite a lot* are averaged across the four categories, many more full intervention teachers selected *a lot* and more partial intervention selected *quite a lot*; as shown in Figure 18. Again, it appears that the cluster meeting professional development the teachers took part in appears to have had an influence.

Figure 18 Difference in responses between the two intervention groups

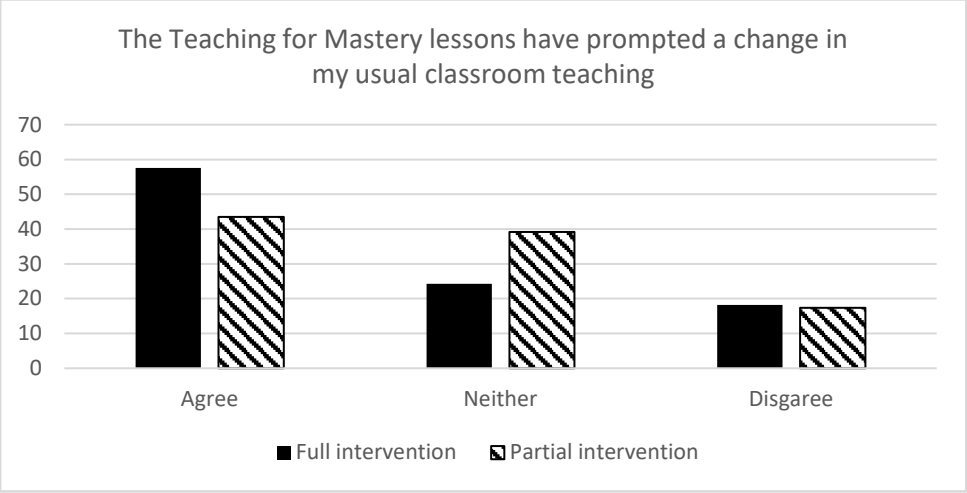


When teachers were asked how well they understood how to adapt their teaching so that it becomes aligned with Teaching for Mastery, over 51.5% of the full intervention teachers reported they understood very well, and 42% reported they understood moderately well. For the partial intervention teachers, on the other hand, the results were 40% and 60% respectively. This suggests that the extra professional development the full intervention teachers participated in may have provided them with more confidence than the partial intervention teachers.

Trial Teachers' change in practice (intermediate outcomes)

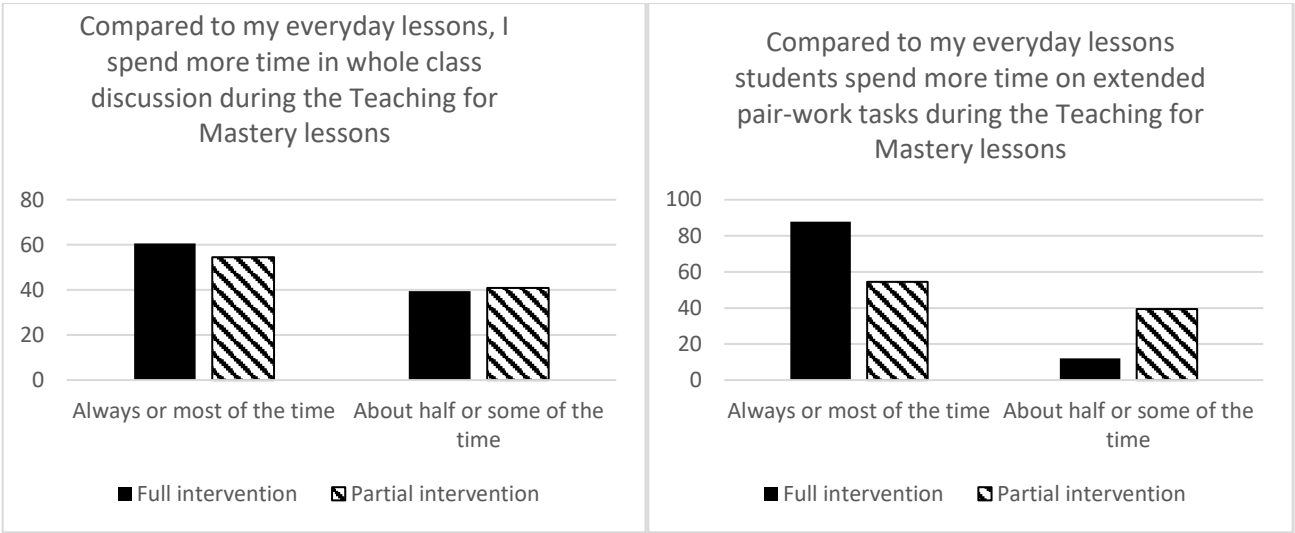
Teachers were asked to what extent they agree with a set of statements related to their use of the Teaching for Mastery lessons. As Figure 19 demonstrates, over half the full intervention teachers agreed that the use of the Teaching for Mastery lessons had prompted a change in their practice, and just under half the partial intervention teachers agreed. Fewer than 20% of the respondents in each group disagreed.

Figure 19 Trial Teachers' perception of change in practice



Specifically, teachers reported two main differences between their everyday teaching and their Teaching for Mastery lessons. First, they spent more time in whole class discussions in the Teaching for Mastery lessons, as Figure 20 shows, with over half of each group reporting that this happened always or most of the time, and a very small difference between the full intervention and the partial intervention group.

Figure 20 Differences between Trial Teachers' everyday teaching and TfM lessons



There is also evidence from the case studies (interviews and observations) that teachers had changed their practice, beyond teaching the lessons, over the course of the intervention.
One teacher, for example, explained that he aimed to teach for understanding in an early interview. In a later interview, he reported that “I'm trying to plan all my lessons in line with mastery as best I can.” (TT14_Int 3). He explained that, in his lessons, he began by exploring what the students already knew.

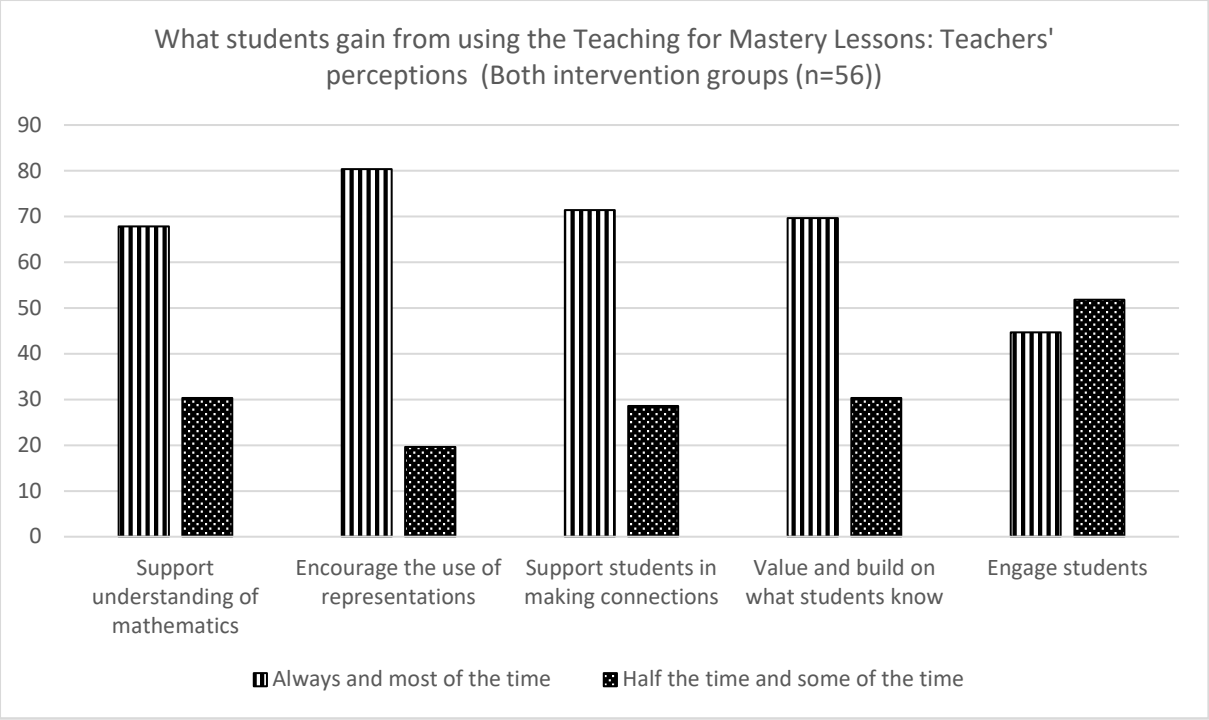
Another teacher talked about focusing on one mathematical idea and developing deep understanding of that idea, further explaining that this is what she was doing in her teaching:

> *What I like about the mastery resources that we've started with one [idea]. And yes, we've stretched out in the end but that key idea is what builds the depth of understanding in the main task, and I think that's what I now want to go and address again in my own practice.”TT157_int 2*

Student engagement (intermediate outcomes)

The Trial Teachers were asked about how the Teaching for Mastery lessons supported their students’ learning. Their responses indicate that they found that the lessons encouraged understanding, used a range of representations and made connections and built on students’ prior knowledge, with most respondents suggesting that this was the case all the time or most of the time (Figure 21). About half the respondents reported that their students found the lessons engaging.

Figure 21 Trial Teachers' perception of student learning



Researcher observations of lessons further confirmed that the students found the lessons engaging, with a report on one lesson stating:

> *It seems that all students were engaged. However, in the pair I observed, once they had placed all the cards, they did not fill in the rest of the table and they just sat. They seemed to be happy to sit. One student (G6) spent a lot of time on her phone, having to hotspot with G1's phone. However, she did all the work and was engaged. (TT67_L3obs)*

The case study teachers also agreed that the students were engaged. One teacher suggested that that they were more engaged than in other lessons, saying

> *They all engaged and were really interested in it, right? I need to bring you into one of the other lessons where they don't engage and they're not interested. (TT12_Int2)*

Another Trial Teacher explained in an interview how pleased she was with the engagement of one particular group, saying:

> *[I was] pleased with how they engaged, particularly in my childcare group. I was really pleased with how they had engaged, yeah. And I think they were sort of in it got to the point where I think where they were enjoying the challenge because they knew that they were solving some of them. (TT_67_Int 3)*

Table 28 below shows the percentages of students in both intervention groups selecting either 'always or often' or 'sometimes or rarely' in areas related to the mastery lessons. In the post-intervention surveys, there is a small increase over the pre-intervention surveys of always and often.

Table 28 Students' perception of the TfM lessons

	Always and often (PRE)	Sometimes and rarely (PRE)	Always and often (POST)	Sometimes and rarely (POST)
The teacher tells us what value the lesson topic has for future use.	64	36	67	33
The teacher explains the underlying structure of the maths	71	29	74	26
The teacher asks us what we already know and builds on this	82	18	83	17
The teacher explains connections between different topics	79	21	80	20
The teacher focuses on developing our understanding of both key ideas and fluency	84	16	84	16

Cost evaluation results

Delivery of Maths Mastery in FE cost approximately £4107.50 per setting in the 2022/23 academic year in which it was delivered. The start-up related costs were the most significant and occurred within the first year being associated with the training for Lead and trial teachers, lesson study preparation time, cluster meeting supply cover, and travel. In Year 2 and Year 3 of implementation, these costs would be expected to reduce with 2 additional lessons being prepared, delivered, and discussed at cluster meetings.

To calculate the total cost per student over 3 years, we assumed the intervention would be delivered to a cohort of 30 Mathematics re-sit students taught by the trial teacher. For larger FE settings the cost per student is likely to be significantly lower. We have assumed 90 students receiving the intervention, although given that the condition of funding can require students to re-sit in the two years to the age of 18. Based on these assumptions, the total cost per student per FE setting over three years is £50.02. In the materials only arm, there were no lesson study cluster meetings and so the total cost per student is reduced to £18.92.

Table 29 Cost of delivering Maths Mastery in FE per setting for full intervention group

Item	Type	Cost	Total cost over three years	Total cost per student per year over 3 years
Lead Teacher Training (3x1 day)	Start up (Year 1 only)	£900	£900	£10
Trial Teacher PD cover to attend training	Start up (Year 1 only)	£900	£900	£10
Lesson Study: preparation time and wrap around	Start Up (Year 1) with costs reducing in Year 2 and Year 3.	£600	£1080	£12
Lesson Study Cluster Meetings	Start Up (Year 1) with costs reducing in Year 2 and Year 3.	£1500	£1200	£13.33
Lesson Study Travel	Start Up (Year 1) with costs reducing in Year 2 and Year 3.	£200	£400	£4.44
Photocopying worksheets (estimate)	Recurring cost	£7.50	£22.5	£0.25

Table 30 Cost of delivering Maths Mastery in FE per setting for partial intervention group

Item	Type	Cost	Total cost over three years	Total cost per student per year over 3 years
Trial Teacher PD cover to attend training	Start up (Year 1 only)	£600	£600	£6.67
Lesson Study: preparation time and wrap around	Recurring cost Per Year	£600	£1080	£12
Lesson Study Cluster Meetings	Recurring cost Per Year	£0	£0	£0
Lesson Study Travel	Recurring cost Per Year	£0	£0	£0
Photocopying worksheets (estimate)	Recurring cost	£7.50	£22.5	£0.25

Conclusion

Table 31 Key conclusions

Key Conclusions
1. GCSE resit students taught by teachers in the full Mastery Teaching intervention made one month additional progress in mathematics learning compared to students in other (business as usual) colleges.
2. Students having had Free School Meals prior to college, and taught by teachers participating in the full Mastery Teaching intervention, made two months additional progress in mathematics learning compared to students in other (business as usual) colleges.
3. Teachers in both intervention groups report that taking part in the PD intervention programme and teaching the exemplary Teaching for Mastery lessons as: ■ being effective as an introduction to the principles of Teaching for Mastery ■ leading to their improved understanding of how to implement Teaching for Mastery in their practice ■ leading to changes in their teaching practice during the programme and high levels of intended change in teaching practice (in subsequent years) ■ resulting in improved student engagement and understanding.
4. Compliance was generally fair, with close to two-thirds of settings in both arms of the intervention teaching lessons 1-4. Compliance decreased for both arms in the teaching of lesson 5 with over half of teachers in the full intervention, but only just over two-fifths of the partial intervention teachers Teaching the final lesson. There was high fidelity in terms of the lesson aims and design in the teaching of the sample lessons and teachers reported trying to implement the Teaching for Mastery approaches in their other lessons.

Impact evaluation and IPE integration

The intervention logic model and outcomes

In general, the intervention was delivered as described in the logic model. However, Lead Teacher and Trial Teacher professional development sessions were held remotely rather than face to face because of background covid causing difficulties for teachers in colleges due to illness and restrictions imposed by colleges due to infection rates (particularly during the winter months). Nonetheless that professional development for both Lead and Trial Teachers was appreciated and effective at introducing the Teaching for Mastery approach. The teachers self-report that the professional development and the exemplary Teaching for Mastery lessons were useful in developing their practice in line with the designed intentions informed by the Key Principles. The

professional development was relatively limited for the Trial Teachers but the lesson plans gave much guidance and prompted reflections on how the Key Principles could be used to guide Teaching for Mastery. This may be of particular significance in the case of teachers in the partial intervention who for most of the time worked at distance and on their own. The teachers in the full intervention had many more opportunities to engage with the teaching approach due to the cluster meetings in which they took part on five occasions spaced throughout the period. These meetings followed a Lesson Study approach involving clusters of teachers coming together to observe one of the group teach the lesson, which all had had an opportunity to teach to their class(es) prior to the meeting. Following the observation there was a post-lesson discussion guided by a Lead Teacher and two 'research' questions associated with each lesson. When designing the intervention, it was hypothesised that this model would be more effective in ensuring teacher engagement with and understanding, of the Teaching for Mastery approach.

The trial outcomes support the intentions of the logic model: that the Teaching for Mastery approach will lead to improved outcomes for students. Although, the research was carried out at a time of disturbance due to the Covid-19 pandemic, we detect impact in terms of GCSE scores, albeit at a low level. This translates to one additional month of learning for students taught by teachers in the full intervention and this increases to two months for students who have been in receipt of free school meals in the last six years if taught by these teachers (when contrasted with students taught in business-as-usual groups). There is a smaller effect size for students taught by teachers in the partial intervention: it is not large enough to indicate any additional months of learning.

Interpretation

The design of the intervention appears to have the desired effect: that is, a change in teachers' practice leading to improved learning outcomes for students. The intervention that incorporates an element of lesson study appears to be more effective as it is students taught by teachers who have taken part in this element of intervention for whom the effect is greatest. This indicates that this model of intervention provides teachers with a more effective experience of engaging with, and incorporating into their practice, Teaching for Mastery. This is perhaps not surprising as Lesson Study has attracted the attention of educators around the world – as a professional learning experience it connects directly with the core business of teachers, that of teaching in classrooms. It is not by chance that it meets all the key features of good quality professional development that research points us to. Research suggests that high quality professional learning activity is:

- Experiential: stimulating & drawing on teachers' experiences.
- Sustained: cycles of planning, predicting, enactment & reflection.
- Grounded: practical, well-resourced; related to context & culture.
- Safe: teachers able to speak their minds, permission to take risks.

- Collaborative: involving networks of teachers & administrators.
- Informed: by outside expertise and research.
- Provocative: involving both pressure and support.
- Focused: attentive to the development of the mathematics itself.[11]

Importantly, Lesson Study can ensure learning of both the individual and the collective/community (Wake, 2003).

There are, however, not many studies that can substantially point to its impact on student learning, even though advocates are convinced of its efficacy. One RCT that did find an impact on student learning was for just a small part of the mathematics curriculum (Lewis & Perry, 2017). We are unaware of any studies that report on the approach being used across the curriculum elsewhere,

Of particular interest is how the impact on student outcomes is magnified for students who come from the most deprived backgrounds (measured by their previous uptake of free school meals). These students are a substantial proportion of the cohort. As we know, students from deprived backgrounds do significantly less well at GCSE (see for example de Moira et al (2020)). A question to ask is, why do these students benefit most from the Teaching for Mastery approach? It is clear that the expectations of the approach, as summarised in the five key principles, are such that there is an emphasis on paying attention to prior learning and providing for collaborative building of understanding in lessons as well as ensuring that the classroom becomes a safe (and social) space where all thinking is valued. Observations of lessons provide evidence that in general this was achieved, and this stands in contrast to students' prior experience of mathematics lessons in schools where there is possibly a greater range of maths attainment in a class and a more didactical approach to teaching. Outcomes suggest that the change in classroom expectations and experience are most beneficial to students who come from deprived backgrounds.

Limitations, lessons learned and future research directions.

In a number of places throughout this report we have highlighted that the research was undertaken at a time when Covid-19 continued to impact on education widely and importantly for the research on colleges and teachers' abilities to commit to, and engage with, the research. This undoubtedly led to high levels of attrition of teachers from those recruited. Particularly, from December through February teachers struggled to ensure that all classes in colleges were able to be covered due to, at that time, their own illness and/or that of colleagues. Their workload was often unbearable and for

[11] See, for example, Guskey & Yoon (2009); Joubert and Sutherland (2009); Villegas-Reimers (2003)

that reason a number had to drop out of the research. Likewise, students were badly affected and their levels of attendance, often low at the best of times, were lowered. The students in the Trial Teachers' classes had also had a poor experience of learning in the two years before the Trials due to covid with substantial periods of lockdowns and learning remotely. In summary, the RCT was not carried out at the best time to give it a fair run.

However, the impact evaluation does point to an improvement in GCSE scores, albeit small, and for some students: those taught by teachers in the full intervention. This impact is most for the most deprived students (as measured by Free School Meals). We do caution about taking too much from the results we report here, but suggest that the research does point to the potential to further investigate the full intervention model, perhaps with a delayed post-test to capture outcomes when teachers have had an opportunity to embed what they learn during the intervention year more substantially in their practice.

The research team plan to publish several academic journal papers summarising and extending the findings of this evaluation report.

References

Angrosino, M. V., & Mays de Perez, K. A. (2000). Rethinking observation: From method to context. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd Ed.) (pp. 673-702). Sage.

Association of Colleges. (2020). *College Key Facts 2019-20*. London: A.o.C.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

BEIS. (2017). *Industrial Strategy: building a Britain fit for the future*. London: HMSO.

Boylan, M., Maxwell, B., Wolstenholme, C., Jay, T., & Demack, S. (2018). The mathemematics teacher exchange and 'mastery' in England: The evidence for the efficacy of component practices. *Education Sciences*, *8*(4), 202. https://doi.org/10.3390/educsci8040202

Bulus, M., Dong, N., Kelcey, B., & Spybrook, J. (2021). *PowerUpR: Power analysis tools for multilevel randomized experiments*. R package version 1.1.0, https://CRAN.R-project.org/package=PowerUpR

Confederation of British Industry. (2015). *Inspiring Growth: CBI/Pearson Education and Skills Survey 2015.* Pearson Education: Harlow.

Coppock, A., Copper, J., Fultz, N., & Blair, G. (2023). *randomizr: Easy-to-use tools for common forms of random assignment and sampling*. R package version 0.24.0, https://cran.r-project.org/package=randomizr

DfE. (2019). *Revised A level and other 16-18 results in England, 2017/2018*, London: DfE, https://www.gov.uk/government/statistics/a-level-and-other-16-to-18-results-2017-to-2018-revised

Gelman, A., & Rubin, D. (1992), Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457-511.

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan*, *90*(7), 495-500.

Higton, J., Archer, R., Dalby, D., Robinson, S., Birkin, G., Stutz, A., Smith, R., & Duckworth, V. (2017). *Effective practice in the delivery and teaching of English and mathematics to 16-18 year olds*. Department for Education. https://www.gov.uk/government/publications/english-and-maths-for-16-to-18-year-olds-effective-teaching

Joubert, M. V., & Sutherland, R. J. (2009). *A perspective on the literature: CPD for teachers of mathematics.* National Centre for Excellence in the Teaching of Mathematics.

Kawulich, B. B. (2005). Participant observation as a data collection method. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, *6*(2).

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270-280. DOI: 10.1177/2515245918771304

Lewis, C., & Perry, R. (2017). Lesson study to scale up research-based knowledge: A randomized, controlled trial of fractions learning. *Journal for Research in Mathematics Education*, *48*(3), 261-299. DOI: 10.5951/jresematheduc.48.3.0261

Makowski, D., Ben-Shachar, M., Annabel Chen, S. H., & Ludecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology*, *10*, 2767. DOI: 10.3389/fpsyg.2019.02767

de Moira, A. P., Meadows, M., & Baird, J. (2020). The SES equity gap and the reform from modular to linear GCSE mathematics. *British Educational Research Journal*, *46*(2), 421–436. DOI: 10.1002/berj.3585.

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

Quartagno, M., & Carpenter, J. (2022). *jomo: A package for multilevel joint modelling multiple imputation*. https://CRAN.R-project.org/package=jomo

Robey, C., & Jones, E. (2015). *Engaging learners in GCSE English and maths*. NIACE. http://shop.niace.org.uk/engaging-learners-gcse-maths-english.html

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-and-government-response

The Research Base. (2014). *Effective Practices in Post-16 Vocational Maths*. London: The Education and Training Foundation.

Villegas-Reimers, E. (2003). *Teacher professional development: An international review of the literature.* International Institute of Educational Planning, UNESCO: Paris.

Wake, G., Hodgen, J., Adkins, M., Ainsworth, S., & Evans, S. (2022). *Young Enterprise: Mathematics in Context, a two-armed cluster randomised trial: Evaluation report*. Education Endowment Foundation. https://d2tic4wvo1iusb.cloudfront.net/documents/projects/YEMiC_Evaluation_Report.pdf

Wake, G. (2023), Designing lesson study for individual and collective learning: networking theoretical perspectives. *International Journal for Lesson and Learning Studies*, *12*(1), 7-20. DOI: 10.1108/IJLLS-08-2022-0111

Williams, J. (2012). Use and exchange value in mathematics education: contemporary CHAT meets Bourdieu's sociology. *Educational Studies in Mathematics*, *80*(1/2), 57-72. DOI: 10.1007/s10649-011-9362-x

Wolf, A. (2011). *Review of vocational education – The Wolf report*. GOV.UK. https://www.gov.uk/government/publications/review-of-vocational-education-the-wolf-report

Tables of tables

Table of figures

Appendix A

About the lessons

The lessons were all designed to exemplify the five 'Key Principles' for Teaching for Mastery in FE:

- Develop an understanding of mathematical structure
- Value and build on students' prior learning
- Prioritise curriculum coherence and connections
- Develop both understanding and fluency in mathematics
- Develop a collaborative culture in which everyone believes everyone can succeed

Each lesson focuses in particular on one of the Key Principles and is accompanied by two research questions, generally designed to highlight the Key principle in focus and aiming to focus the teachers' thinking on this Key Principle.

The design of the lessons is guided by a set of design features which take into account both the mathematics to be taught and ways of working in the classroom.

In terms of the mathematics, the lessons

- address fundamental mathematical ideas/concepts,
- highlight mathematical structure (e.g., by using context, representations, variation),
- foreground common misconceptions through activity that provokes cognitive conflict,
- provide for a range of likely approaches, and
- connect different areas of mathematics where possible.

Lessons include time and opportunity for exploring mathematics and discussing different approaches. This deliberate slowing down is designed to encourage students to communicate using mathematical language and to critique their own and others' approaches. The closing discussion is more than the solution(s) to the student tasks; in it, the teacher works with student thinking and summarises the class collaborative thinking. The model is one that contrasts with that of direct instruction, rather it promotes collaborative sense making and construction of understanding.

Lesson plans provided the teacher with a high level of support about how the Key Principles can be put into practice. Teachers were also given PowerPoint presentations and any other resources such as student task sheets that were needed.

The table below lists the titles, key objectives, Key Principle in focus and research questions for each lesson.

Table 32 Titles, key objectives, Key Principle in focus and research questions for each lesson

Lesson Title	Objectives	Teaching for Mastery Key Principle focus and research questions
Lesson A: Factors and multiples	■ Find factors and multiples of numbers less than 100 ■ Find common factors and common multiples of two numbers ■ Solve problems involving HCF and LCM ■ Write a number as a product of prime factors ■ Understand how to use representations to provide insight into solving problems	Key Principle 1: Develop an understanding of mathematical structure Research questions *Pedagogic focus* How does the teacher use the student tasks to help develop understanding of mathematical structure? *Maths focus* In what ways do students use representations to consider mathematical structure to develop their mathematical understanding of highest common factor and lowest common multiple?
Lesson B: Area and volume	■ Understand the effects on area and volume of scaling one or both dimensions of a rectangle and one, two or all three dimensions of a cuboid ■ Understand and apply conservation of area and volume ■ Use relationships between similar figures to determine areas and volumes	Key Principle 2: Value and build on students' prior learning Research questions *Pedagogic focus* How is the lesson developed and brought to a close in ways that values and builds upon what students already know? *Maths focus* What evidence do you observe of students' prior learning about area and volume and how do they work with or modify this?
Lesson 1: Multiplicative reasoning	■ Understand the multiplicative relationship between two quantities (non-calculator)	Key Principle 2: Value and build on students' prior learning

	■ Solve multi-step currency or unit conversions problems (calculator) ■ Understand how to use representations to provide insight into solving problems	Research questions *Pedagogic focus* How is the lesson developed and brought to a close in ways that values and builds upon what students already know? *Maths focus* What evidence do you observe of students' prior learning about multiplicative reasoning and how do they work with or modify this?
Lesson 2: Ratio and fractions	■ Write a ratio as a fraction and vice versa ■ Divide a given quantity into parts in a given part : part or part : whole ratio ■ Solve multi-step problems using ratios and fractions ■ Use representations to provide insight into solving problems	Key Principle 3: Prioritise curriculum coherence and connections Research questions *Pedagogic focus* How does the teacher use class discussion to emphasise connections in mathematics? *Maths focus* What evidence do you observe of students using representations to develop understanding of, and connections between, fractions and ratios.
Lesson 3: Basic algebra	■ Use correct algebraic notation ■ Expand brackets ■ Factorise algebraic expressions ■ Understand multiplicative algebraic structure using an area representation	Key Principle 1: Develop an understanding of mathematical structure Research questions *Pedagogic focus* How does the teacher promote understanding of mathematical structure? *Maths focus*

		In what ways do students use representations to consider mathematical structure to develop their mathematical understanding?
Lesson 4: Algebraic thinking	■ Represent contextual problems mathematically Use diagrams to represent mathematical structure ■ Determine the value of an unknown in a problem ■ Solve problems involving angles	Key Principle 5: Develop a collaborative culture in which everyone believes everyone can succeed Research questions *Pedagogic focus* In what ways does the teacher develop and bring the lesson to a close to support a culture where everyone believes everyone can succeed? *Maths focus* In what ways do students use representations to access the structure of mathematical problems?
Lesson 5: Percentage change and best buys	■ Become fluent at working with percentage change ■ Determine the best deal following a percentage change ■ Understand different approaches to solving multi-step percentage problems ■ Use representations to provide insight when solving problems	Key Principle 4: Develop both understanding and fluency in mathematics Research questions *Pedagogic focus* In what ways do teachers use diagnostic and formative assessment to support students in becoming fluent? *Maths focus* In what ways do students develop their understanding and fluency?

Appendix B

About the professional development

Lead Teachers PD: description and objectives

The professional development for the Lead Teachers aimed to develop their understanding of:

- Lesson study
- The design of the seven lessons
- Running cluster meetings

The model consisted of three professional development days specifically for Lead Teachers, organised and co-ordinated by the University of Nottingham. The professional development provided by the UoN modelled the ways in which it was expected that LTs would work with teachers in their cluster groups,

Each LT professional development day was different, but all included a review of the key principles for Teaching for Mastery in FE, detailed introductions to the lessons for the upcoming windows and some discussion of lesson study and cluster meetings. The schedule for the second Lead teacher professional development day (23rd June 2021), below, provides an example.

Time	Content
9:30 – 9:45	Welcome and introductions Updates
9:45 – 10:10	Teaching for Mastery
10:10 – 10:30	Lesson study part 1
10:30 – 10:50	Break
10:50 – 11:50	Lesson 1: multiplicative reasoning (hotel costs) Breakout rooms
11:50 – 11:55	Comfort break
11:55 12:10	Trials, recruitment, case studies, data collection
12:10 – 12:55	Lesson 2 ratio (sharing baguettes)
12:55 – 13:45	Lunch break
13:45 – 14:00	Lesson study part 2
14:00 – 14:10	LTs and cluster groups
14:10 – 14:15	Little break
14:15 – 15:00	Lesson 3 algebra
15:00 – 15:15	Break
15:15 – 15:45	Next steps Questions Evaluation

Lead Teachers PD: evaluation

The LT professional development was evaluated in two ways:

- A post-session survey for each of the three professional development days (in March 2021, June 2021, January 2022)
- Interviews with LTs of case study teachers (in March 2021 and June 2021)

The post-session survey was administered electronically via a personalised link sent to each LT.

The questions on the survey are provided here:

Session evaluation: Lead Teacher PD

1. At this point in the project, how well do feel you understand:
    a. The role of a Lead Teacher for the Teaching for Mastery project
    b. Lesson study
    c. Teaching for Mastery
    d. Running a cluster meeting

Response options:
Very well *Moderately well* *Not very well* *Not at all*

2. To what extent has your understanding of the following changed as a result of the PD session today?
    a. The role of a Lead Teacher for the Teaching for Mastery project
    b. Lesson study
    c. Teaching for Mastery
    d. Running a cluster meeting

Response options:
Major increase Moderate increase Slight increase No increase

3. Please indicate how engaged you were with the presentations and activities today.

Response options:
Very engaged Moderately engaged Mildly engaged Not at all engaged

4. Please write a short statement to explain your answer to Question 3.

5. Please add any further explanation of your responses, comments about the PD session and/or your preparation to be a Lead Teacher (optional).

The relevant interview questions are given here:

Q	Main questions and prompts	RQ
1	The objectives of the PD days were to develop your understanding of three things: a) the lesson study process; b) the lead teacher role; c) teaching for Mastery in FE. I am interested in how confident you were before the PD sessions about these three areas and how your understanding developed.	1.3/5.1

	Please could we first talk about the Mastery approach and then the other aspects.	
2	What parts of the first PD sessions do you think have been the most useful to prepare and support you for facilitating cluster meetings?	5.1
3	How engaging did you find the materials provided, the presentations, discussions and other parts of the PD sessions, and to what extent did you engage in the PD activities?	3.1
4	How well prepared overall do you feel as a result of the PD for leading the cluster meetings? What else might be useful?	5.1
5	Do you have any other comments or suggestions to make about the PD and your preparation for being a lead teacher?	3.1/5.1

Supports for LTs

LTs were provided with:

- A LT guide which included sections on Teaching for Mastery in FE, Collaborative learning, the nationals trials, lesson study and cluster meetings. This can be found here.
- Resources to support them in leading each cluster meeting, which included, at least, written guidance, a PowerPoint and an observation form for the lesson study. As an example, links to the resources for Cluster Meeting 2 can be found here:
    - Written guidance
    - PowerPoint
    - Lesson observation form

Trial Teachers PD: description and objectives

All teachers in the full- and partial-intervention groups took part in the equivalent of one day's professional development in each of October 2021 (Mastery 1), January 2022 (Mastery 2) and June 2022 (Mastery 3). Overall, the professional development aimed to develop teachers' understanding of:

- The Teaching for Mastery Key Principles for FE
- How the sample Teaching for Mastery lessons were designed and how they should be used
- How they might adapt their own teaching so that it becomes more aligned with Teaching for Mastery approaches

The teachers in Group 1 were grouped into 'clusters' of about four teachers, based on their physical location. A Lead Teacher was assigned to each cluster group, also based mainly on location. The Lead Teacher ran a cluster meeting in each window. At each cluster meeting the group observed a colleague teaching the lesson for that window using an adapted lesson study approach, led by the

Lead Teacher. The Lead Teacher also introduced the teachers to the lesson to be taught in the following window. An additional cluster meeting took place before the first window, in which the Lead Teacher introduced Lesson 1.

The cluster meeting professional development aimed to improve or develop teachers' capacity to:

- Analyse teaching and learning
- Reflect on their own practice

Trial Teachers PD: evaluation

The TT professional development was evaluated in two ways:

- A post-session survey for each of the three professional development days (in October 2021, June January 2022, June 2022)
- Interviews with case study teachers (in October 2021, December 2021 and March 2022)

Session evaluation: Trial Teacher PD (Mastery 1 and Mastery 2)

The evaluation will follow the morning PD session on **October 15th.** The questions will be the same for trial teachers in **Groups 1 and 2.**

1. At this point in the project, how well do feel you understand:
   a. the teaching for Mastery in FE approach overall?
   b. the ways in which student tasks and representations are used to develop an understanding of mathematical structure?
   c. the ways in which teachers value and build on the prior learning of students?
   d. your role as a Group 1 or 2 trial teacher?

Response options:
Very well *Moderately well* *Not very well* *Not at all*

2. To what extent has your understanding of the following changed as a result of the PD sessions Mastery 1 and Mastery 2?

   a. the teaching for Mastery in FE approach overall?

Response options:
Major increase *Moderate increase* *Slight increase* *No increase*

Please explain your answer, stating which parts of the PD sessions were most helpful, if any.

   b. the way in which student tasks and representations are used to develop an understanding of mathematical structure?

Response options:
Major increase *Moderate increase* *Slight increase* *No increase*

Please explain your answer, stating which parts of the PD sessions were most helpful, if any.

c. the ways in which teachers value and build on the prior learning of students?

Response options:
Major increase *Moderate increase* *Slight increase* *No increase*

Please explain your answer, stating which parts of the PD sessions were most helpful, if any.

d. your role as a Group 1 or Group 2 trial teacher?

Response options:
Major increase *Moderate increase* *Slight increase* *No increase*

Please explain your answer, stating which parts of the PD sessions were most helpful, if any.

3. How engaged were you with the Mastery 1 and Mastery 2 PD presentations and activities?

Response options:
Very engaged *Moderately engaged* *Mildly engaged* *Not at all engaged*

4. Please explain your answer to Question 4.

5. Overall, how prepared do you feel to use teaching for Mastery approaches in your classroom?

Response options:
Very well *Moderately well* *Not very well* *Not at all.*

6. Please explain your answer to Question 5.

7. Please add any further explanation of your responses, comments about the Mastery 1 and 2 PD sessions (optional) or your preparation to be a Group 1 or 2 trial teacher.

Support for Trial Teachers

TTs were provided with:

- A TT handbook, which explained what trial teachers in the different arms of the trial needed to do, and when. This can be found here.
- A data collection handbook, which explained in detail what data needed to be provided, when and how. This can be found here.
- Resources to support teaching Lessons 1 to 5, and Lessons A and B, which can be found here.

Appendix C

GCSE Mastery scale

Setter	Paper	Items
Edexcel	1F	Q14, Q18, Q19, Q25, Q29
Edexcel	2F	Q16, Q19, Q21b,c, Q24a, Q25, Q26, Q27
Edexcel	3F	Q1, Q11, Q17a,b,c, Q24, Q25, Q27, Q28, Q29, Q30
AQA	1F	Q2, Q7, Q9, Q10b, Q22, Q26, Q28
AQA	2F	Q13, Q14, Q15, Q17b,c, Q18a,b, Q23, Q24, Q25b, Q26, Q28
AQA	3F	Q4, Q8, Q12a,b,c, Q13, Q19a,b, Q22, Q27, Q28
WJEC	1F	Q6a,b, Q7a,b, Q9a,b, Q12b, Q13, Q14a,b, Q18b, Q19, Q21b, Q22a,b, Q25a,bi, Q26a
WJEC	2F	Q2a, Q4b, Q6a,b, Q9a,b,c, Q10a,b,d,e, Q13, Q19a,b, Q20
OCR	1	Q3b, Q7a,b, Q13b, Q16, Q19, Q21f, Q24a
OCR	2	Q2d, Q5, Q6b,c, Q7a, Q9, Q10, Q11a, Q16a,b, Q22
OCR	3	Q6, Q7a,b, Q9, Q11, Q12, Q18a, Q19a,b, Q22, Q24a, Q25