



Department
for Education

Adult numeracy randomised controlled trials: Adapted mastery approach

**Cluster pilot RCT and implementation and
process evaluation**

May 2026

**Authors: Dimitris Vallis, Gabrielle McGannon,
Megan Liskey, Aida Weheba, Johnny Runge, and
Michael Sanders, The Policy Institute at King's
College London**



Government
Social Research

Contents

List of figures	4
List of tables	5
Acknowledgements	7
Executive summary	8
Key findings	8
Background and rationale	9
Intervention	10
Methodology and process	10
Impact findings	12
Implementation and process findings	13
Conclusions and recommendations	13
1. Introduction	16
1.2. Intervention description: theory of change	18
1.3. Intervention description	21
1.4. Evaluation overview	24
1.5. Research questions	24
1.6. Reading the report	26
2. Methods	27
2.1. Impact evaluation	27
2.2. Implementation and process evaluation	37
3. Impact evaluation results	45
3.1. Participants and attrition	45
3.2. Outcomes and analysis	52
IPE results	62
3.3. Compliance	63
3.4. Attrition	65
3.5. Fidelity	66
3.6. Feasibility	69
3.7. Readiness for trial	76
3.8. Perceived outcomes	77

4. Conclusions	87
4.1. Lessons learned for future delivery	89
4.2. Considerations for future research	89
5. References	91
Appendix 1: Glossary	93
Appendix 2: Detailed intervention description	105
Appendix 3: Costs	115
Appendix 4: IPE research sub-questions and tools	117
Appendix 5: Code	142
Appendix 6: Additional tables	161

List of figures

Figure 1: Gateway trial design	29
Figure 2: Provider attrition map.....	46
Figure 3: Distribution of confidence in maths	53
Figure 4: Distribution of confidence in maths (composite)	54
Figure 5: Attendance (average cumulative)	56
Figure 6: Attendance (by session)	56
Figure 7. Overall, how satisfied or dissatisfied were you with the course?	73
Figure 8. Overall, how engaging did you find the course?	75
Figure 9: What impact, if any, do you think the course has had on your.....	80
Figure 10. To what extent, if any, did your experience of the course influence your decision to do more study or training?	82

List of tables

Table 1: Theory of Change	19
Table 2: Trial design table	27
Table 3: Number of responses and response rates for learner surveys	33
Table 4: Overview of IPE interviews	39
Table 5: Provider region	47
Table 6: Baseline outcome summary statistics	47
Table 7: Sample proportions by baseline category	49
Table 8: Balance of baseline characteristics	51
Table 9: Attrition – secondary outcomes	52
Table 10: Results – primary outcome (attainment)	53
Table 11: Results – secondary outcomes (confidence in maths)	54
Table 12: Results for individual questions in composite confidence in maths measure	55
Table 13: Results – CACE	56
Table 14: Missing data – baseline covariates	57
Table 15: Results – MICE	58
Table 16: Sub-group effects of treatment interactions	59
Table 17: Results – robustness check	59
Table 18: Results – conditional effect sizes	60
Table 19: Results – unconditional effect sizes	60
Table 20: ICC conditional estimates	61
Table 21: ICC unconditional estimates	61
Table 22: Attendance rates across sessions in the treatment group	64
Table 23: Would you say that the maths covered on your course was...	74
Table 24: Overall, how confident do you feel working with numbers in everyday life?	78

Table 25: How confident do you feel about doing the following things in everyday life?	79
Table 26: What impact, if any, do you think the course has had on your...	81
Table 27: Are you planning to do any more courses or training?	81
Table 28: Overview of lessons	105
Table 29: Attrition predictiveness	161

Acknowledgements

We would like to thank all the adult education institutions who participated in the trials, in particular the staff members and learners who participated in interviews about their experiences of the Adapted Mastery Approach (AMA). Their insights were invaluable to the research.

We are also grateful for the managed service supplier (MSS), Etio, and the product developers, the Education and Training Foundation (ETF), for their participation in interviews and their contribution to intervention development, marketing and trial recruitment. Thank you to colleagues at the Department for Education (DfE) for their support throughout the design and delivery of the trials, as well as for their feedback on research tools and this report.

Finally, we would like to thank Ipsos UK for managing the consortium and collecting and managing the trial data, as well as our consortium partners at RAND Europe, Learning and Work Institute, and the Institute for Employment Studies.

Executive summary

This report presents findings from a pilot randomised controlled trial (RCT) and implementation and process evaluation (IPE) of the Adapted Mastery Approach (AMA) for Functional Skills Qualification (FSQ) Level 1 maths delivered in adult education settings. This study was funded by the Department for Education (DfE) as part of a programme of Adult Numeracy Trials delivered through the Multiply programme. The primary outcome for the pilot RCT was maths attainment, and the secondary outcome was confidence in maths. The IPE explored the extent to which the intervention was delivered as intended, participant engagement, how realistic and effective the intervention was, perceived outcomes among participants, and barriers and enablers to delivery.

Key findings

The pilot trial found no statistically significant evidence that AMA improved FSQ Level 1 maths pass rates compared to the control group who received FSQ Level 1 maths courses without the AMA intervention. Learners in the treatment group were 0.1 percentage points (ppt) more likely to pass than those in the control group ($p=.995$),¹ with a 51.1% probability of passing in the control group and a 51.2% probability in the treatment group. However, this was not statistically significant at the 95% confidence level.² Surveys and interviews with tutors and learners who participated in AMA found that there were perceived improvements in learners' confidence in maths, willingness to attempt maths, and interest in progressing to further study in maths. However, it should be noted that similar improvements in outcomes were also reported for learners who received the business as usual (BAU) FSQ Level 1 maths course.

When including whether a learner participated fully in the intervention in the analysis, the effect size increased (3.3ppt) but remained not statistically significant ($p=.782$). However, as a pilot trial with a small number of participants, the trial was not sufficiently powered to detect small to moderate effects. This means that modest but potentially meaningful benefits from the intervention were unlikely to have been detected, and the results are not definitive proof of no impact.

¹ P-values indicate the likelihood that observed differences occurred by chance. The smaller the p-value, the stronger the evidence that the intervention made a real difference. Values below 0.05 are considered statistically significant (i.e. unlikely to have occurred by chance).

² When a difference is 'statistically significant at the 95% confidence level', researchers can be 95% certain the difference is real and not just due to chance. The 95% confidence level is widely accepted as the standard threshold for determining statistical significance in research and evaluation.

The study found some evidence that AMA resulted in a small improvement to learners' overall maths confidence. While the main score (based on one question related to confidence working with numbers in everyday life) was not statistically significant ($p=.956$), there was evidence of a small positive effect on the composite maths confidence score (based on 6 different questions related to different aspects of confidence). However, this was only significant at the 90% confidence level, not the conventional 95% confidence level ($p=.066$).

Tutors and learners reported strong engagement and satisfaction with the AMA programme, with learners reporting improvements to their confidence, problem-solving, and ability to apply maths in real-world contexts. Although some elements of AMA were similar to business as usual (BAU) practice, tutors broadly viewed it as a different approach, and awareness of mastery principles and approaches was low among control tutors.

A high number of providers dropped out of the trial (32% of treatment providers and 37% of control providers) for reasons including staff turnover, concerns that AMA did not fit their existing delivery model, problems with recruiting learning for the FSQ Level 1 maths course, and the burden of training and deliver. However, those who remained delivered the programme broadly as intended. Training was completed as planned, lesson plans were followed, and attendance at training and study sessions was high. Findings suggest AMA can, broadly, be realistically and effectively implemented. However, the high dropout rates may limit the extent to which findings reflect the experiences of the full sample of providers, as those who remained in the trial may have been more motivated or better positioned to implement the intervention.

The study identified barriers to implementation around alignment with the curriculum, suitability for different learner groups and online delivery. The time required for tutors to attend lesson study sessions was also a barrier, indicating that some adaptations might be needed for wider rollout.

Background and rationale

The Adult Numeracy Trials were funded through the government's Multiply programme, which ran from April 2022 to March 2025 and provided free numeracy courses for adult learners across England. Funding for the research element was in place until March 2026. The trials aimed to generate robust, high-quality evidence on the impact of specific interventions designed to engage, motivate and teach essential maths skills to adults, and to understand the feasibility, opportunities and challenges of implementing trials within the adult education sector. The aim was to address evidence gaps and support broader efforts to ensure value for money in adult education. The trials were innovative and experimental and some of the first of their kind within the adult education sector. As with all RCTs, the interventions were tested to see what difference they made to adult learner outcomes, through identifying, in a statistically robust way, those which show

measurable impact when compared to a randomised control group. In doing so, the trials aimed not only to establish what works for adult numeracy learning, but also to generate valuable learning about how RCTs can be effectively designed and delivered in this diverse sector. The Maths Mastery approach, developed by the Education and Training Foundation (ETF) as part of the DfE-funded Centres for Excellence in Maths (CfEM), asserts that all individuals can enjoy and excel in maths. The curriculum provides a coherent and detailed sequence of essential content, ensuring sustained progression over time. Lessons are closely linked to prior learning, carefully sequencing steps to build a secure understanding, while various instructional methods are employed, such as whole-class interactive teaching, questioning, short tasks and discussion (CfEM, 2023). Departing from the standard delivery approach could be particularly beneficial for adult learners who have often had negative previous experiences at school. There is a significant body of research showing that mastery approaches can improve maths attainment (Boylan et al., 2018). Previous evidence on mastery interventions has found promising results for primary and secondary pupils, as well as post-16 learners (Jerrim & Vignoles, 2015; Wake et al., 2023).

Intervention

The AMA intervention built on an existing Maths Mastery approach, originally developed in the UK for teaching in primary and secondary schools by the ETF. AMA applied these principles to adult education settings and was underpinned by 5 core principles: developing understanding of mathematical structures, building on prior learning, prioritising curriculum coherence and connections, strengthening fluency and understanding, and fostering a culture in which all learners are expected to succeed. The intervention was designed for adult learners (aged 19+) studying FSQ Level 1 maths courses. This qualification focuses on practical maths skills needed for everyday life and work, signalling to employers that learners possess the foundational level of maths required for many roles. FSQ Level 1 maths courses typically target individuals who require a more applied or vocational approach to learning maths.

Tutors received 10 hours of initial training, structured lesson plans and teaching resources, and ongoing weekly study sessions facilitated by a continuing professional development (CPD) expert from ETF. For learners, the intervention involved 22.5 hours of AMA lessons over either 15 or 30 weeks, which were embedded into the existing FSQ Level 1 curriculum. The control group received BAU FSQ Level 1 maths provision.

Methodology and process

The **primary research question** was:

What is the average difference in FSQ Level 1 maths attainment, measured by pass rate (binary outcome of pass/fail) of adult learners in providers that deliver AMA as part of

their FSQ Level 1 maths course, in comparison to adult learners who do not receive AMA as part of their FSQ Level 1 maths course?

Data for the primary outcome – learner attainment measured as pass or fail in FSQ Level 1 maths - came from the Individualised Learner Record (ILR).³

Due to challenges with data collection, the only secondary outcome included in the final analysis was confidence in maths. Confidence in maths was measured using items from a learner survey completed at the start (baseline) and end (endline) of the learners' FSQ Level 1 maths course. Subgroup analysis was also conducted across demographic characteristics including sex, ethnicity and health/disability status.

The impact evaluation was delivered as a 3-arm cluster pilot RCT, with randomisation at the provider level. The trial was conducted between May 2023 and March 2026, with intervention delivery taking place during the 2024/25 academic year. Results from the pilot trial's third arm (Contextualised Approach) are outlined in a separate report (Newton et al., 2026).

The pilot trial merged 2 previously separate trials (Contextualised Approach and AMA) into a single design known as a gateway trial. Providers were first offered the 3-arm trial (Contextualised Approach; AMA; and control, where learners received FSQ Level 1 maths courses without either intervention). However, if a provider had previously engaged with one of the interventions, or felt so strongly against implementing one that they would otherwise withdraw from the trial altogether, they were offered a 2-arm trial. This alternative excluded the intervention they had previously engaged with or did not want to implement (e.g. Contextualised Approach and control; or AMA and control).

Adult learning providers were recruited from among the 480 further education (FE) providers delivering FSQ Level 1 maths courses. A total of 25 providers were recruited to this trial, 13 providers in the treatment group and 12 in the control group. This fell short of the original target of 15 providers per trial arm, which reduced the reliability of the results. Learners were then recruited from participating providers if they were aged 19 and over and enrolled on a FSQ Level 1 maths qualification during the 2024/25 academic year (470 learners in treatment and 444 learners in control). Additionally, outcome data was not available for all participating learners. This is likely to have weakened the findings, as the analysis had to classify learners without outcome data as having failed the exam.

As a pilot trial with a small number of participants, this trial was not sufficiently powered to detect small to moderate effects. The accompanying implementation and process evaluation (IPE) examined how AMA was implemented in practice, including tutors'

³ The ILR is an ongoing collection of data about learners from training providers in the Further Education (FE) and Skills sector in England.

attendance and engagement with training and support, and the extent to which the AMA model and principles were implemented. It also explored tutors' and learners' experiences, perceived outcomes, and barriers and enablers to delivery, to assess feasibility and readiness for wider rollout. IPE explored participants' experiences through baseline and endline surveys with learners from both treatment and control groups, as well as 105 semi-structured interviews with learners, tutors, and providers. The IPE also analysed monitoring information, including data on learner attendance and tutor session delivery.

Following the trial feasibility assessment in June 2023, the trial protocol was initially registered in June 2024, with a second version registered in September 2024. Recruitment took place from May 2024 to February 2025, with randomisation taking place from June to August 2024. The intervention was delivered from September 2024 to July 2025, with baseline surveys running September 2024 to March 2025. IPE fieldwork ran from November 2024 to August 2025, with endline surveys conducted from November 2024 to September 2025. ILR outcome analysis took place from November 2025 and was finalised by March 2026.

Impact findings

Primary outcome (FSQ Level 1 maths pass rate): The analysis found no evidence that AMA improved pass rates. Learners in the treatment group were 0.1 ppt more likely to pass FSQ Level 1 maths (51.1% likelihood of passing in the control group and 51.2% in the treatment group) but this result was not statistically significant ($p=.995$).

Secondary outcome (confidence in maths): The study found no statistically significant evidence that AMA improved learners' overall confidence in maths for the main score which was based on one question related to confidence in working with numbers in everyday life ($p=.956$). There was, however, evidence of a small positive effect on the composite maths confidence score (based on 6 questions related to different aspects of confidence), but this was only statistically significant at the 90% level, not at the conventional 95% threshold ($p=.066$).

Compliance-adjusted analysis: When including a measure of compliance with the intervention in the analysis, defined as learners attending at least 80% of their FSQ Level 1 maths classes, the estimated impact on pass rates increased slightly to 3.3 ppt ($p=.782$), though it remained not statistically significant.

Subgroup analysis: The trial was not able to detect differential treatment impacts across learner demographic characteristics, including sex, ethnicity and health/disability status.

Implementation and process findings

Providers, tutors and learners generally reported positive experiences with AMA. Engagement with tutor training and delivery was high. While treatment tutors noted some similarities to their usual practice, they also described shifts in teaching practices, mindset, and approaches. Treatment learners were largely satisfied with the quality of teaching, with many reporting increased confidence, greater willingness to attempt maths and problem-solving, and improved ability to apply maths in everyday contexts as well as future study or work. Many reported that the course made them more interested in pursuing further maths courses, particularly due to improved confidence and reduced anxiety about maths. Providers and tutors were broadly supportive of the approach, with many intending to continue or expand AMA delivery. Treatment tutors also valued the professional development benefits, including opportunities to learn new teaching approaches and renewed motivation. Control group tutors and learners also reported positive experiences, but awareness of AMA principles among control tutors was generally low. This suggests minimal contamination; that is, the control group did not adopt elements of the AMA approach, which increases confidence in the overall trial findings.

The study found high compliance with training and delivery requirements among tutors and providers. However, 2 treatment providers also participated in another Adult Numeracy Trial, potentially affecting delivery and dropout (also known as attrition). Attrition following randomisation was relatively high among both treatment providers (32%) and control providers (37%). This was driven by staff turnover, concerns that AMA did not fit existing delivery models, inability to recruit learners for courses, and the burden of training and delivery requirements.

Among the treatment providers who stayed in the trial, fidelity was strong overall; that is, the intervention was delivered largely as intended. Most tutors attended pre-delivery training and lesson study sessions, and largely followed the lesson plans, though many made minor adaptations to meet learners' needs and cover the curriculum. The intervention was considered feasible and practical to deliver, including within the FSQ Level 1 maths course hours. There was high satisfaction among tutors and learners. However, barriers were identified around alignment with the curriculum, suitability for learners at different levels, suitability for online delivery, and the time commitment required for weekly lesson study sessions.

Conclusions and recommendations

The pilot trial found no statistically significant evidence that AMA improved pass rates or maths confidence among learners taking FSQ Level 1 maths courses. However, there was some evidence that learners' confidence in maths (as measured by the composite score) may have increased; further analysis of each individual question within the composite score suggests learners felt more confident in specific everyday uses of

maths. Additionally, the IPE found that the intervention was generally well received. Tutors reported positive changes in their teaching practice, mindset and approaches, while learners described increased confidence, greater willingness to attempt maths and problem-solving, and improved ability to apply maths in everyday situations.

However, several challenges may explain why these positive experiences did not translate into measurable impacts in the quantitative analysis. Tutors highlighted barriers to delivering the intervention as intended, resulting in some deviation from the programme design. The trial also experienced high dropout rates among both treatment and control providers after randomisation, and the number of learners varied substantially between participating providers, both of which reduced the trial's ability to detect impacts and raised concerns about the practical burden of the intervention.

There were also limitations to the study. The pilot trial was only powered to detect large improvements in the outcomes of interest. It is therefore possible that AMA had a modest or small impact that this study was unable to detect. The final sample was even smaller than anticipated, reducing statistical precision. High rates of missing outcome data likely also weakened the observed effects. For the IPE, practical constraints and concerns about research burden meant modifications were made to the design from the original protocol. There were also limitations related to interview sampling and data quality from learner surveys and management information. Therefore, IPE findings should be interpreted with caution as they may not represent the wider population of learners and tutors. However, the IPE provides rich insights into the range and diversity of views and experiences among different participants.

Overall, while the impact evaluation did not find significant impacts, the pilot trial provides useful information about the feasibility of conducting a future RCT on mastery approaches in adult learning settings. The study demonstrates that providers and learners can be successfully recruited, that randomisation at the provider level is feasible, and that key outcome data can be collected. However, for potential future research, adaptations should be made to the intervention to improve retention and fidelity, including reducing the training burden and making the course materials more appropriate for a range of delivery models, such as online delivery.

Any future trial design must also ensure a sufficient sample size for adequate statistical power to detect treatment effects. As the impact in this pilot trial was so small (an approximately 0.1 ppt increase in pass rate), it is not feasible to estimate the sample size required to detect this effect in any future trials. However, based on a 10 ppt difference between treatment and control (a conventional effect size estimate), a trial with the same patterns of variation and attrition rates would require approximately 241 providers across both arms, or 44% of all adult education providers across England. The pre-randomisation take-up rates observed in the pilot trial also support the need to recruit a high number of providers if a future trial were to randomise at the provider level, making this design likely infeasible. Therefore, future research should consider other levels of

randomisation (such as individual, class or tutor), while acknowledging the risks of contamination this could introduce.

Finally, this trial was part of a programme of trials on adult numeracy commissioned by the DfE. Alongside the individual trial reports, DfE has published a programme report on findings related to running RCTs in the adult learning sector, describing the broader learnings for the sector (Mackay et al., 2026).

1. Introduction

This report sets out the findings from an impact evaluation and implementation and process evaluation (IPE) of the Adapted Mastery Approach (AMA) for Functional Skills Qualification (FSQ) Level 1 maths in adult education settings. AMA was a new intervention which aimed to improve adult learners' confidence in maths by adopting a maths mastery approach. The impact evaluation was delivered as a 3-arm, pilot cluster randomised controlled trial (RCT). The AMA group involved 13 adult education providers and 470 learners across England. The control group involved 12 adult education providers and 444 learners. Results from the pilot trial's third arm (Contextualised Approach) have been reported on separately.

1.1. Background and rationale

1.1.1. Background of Multiply

The Adult Numeracy Trials were funded through the government's Multiply programme, which ran from April 2022 to March 2025 and provided free numeracy courses for adult learners across England. Funding for the research element was in place until March 2026. The trials aimed to generate robust, high-quality evidence on the impact of specific interventions designed to engage, motivate and teach essential maths skills to adults, and to understand the feasibility, opportunities and challenges of implementing trials within the adult education sector. The aim was to address evidence gaps and support broader efforts to ensure value for money in adult education. The trials were innovative and experimental and some of the first of their kind within the adult education sector. Details on the other trials conducted can be found at [Adult numeracy randomised controlled trials](#). The overall performance of the Multiply programme was subject to a separate [evaluation](#).

As with all RCTs, the interventions were tested to see what difference they made to adult learner outcomes, through identifying, in a statistically robust way, those which show measurable impact when compared to a randomised control group.

In doing so, the trials aimed not only to establish what works for adult numeracy learning, but also to generate valuable learning about how RCTs can be effectively designed and delivered in this diverse sector.

1.1.2. Existing evidence

Despite a statutory entitlement to free English and maths up to Level 2, participation and achievement in adult maths courses had declined in the decade prior to Multiply (DfE data). This trial sought to contribute evidence on adopting a maths mastery approach in adult education settings, specifically in Functional Skills Qualifications (FSQ) Level 1 maths classes.

The maths mastery approach, developed by the Education Training Foundation (ETF) through the DfE-funded Centres of Excellence in Maths (CfEM) programme, asserts that all individuals can enjoy and excel in mathematics. The curriculum is designed to provide a coherent and detailed sequence of essential content, ensuring sustained progression over time. Lessons are intricately linked to prior learning, carefully sequencing steps to build a secure understanding, and various instructional methods are employed, such as whole-class interactive teaching, questioning, short tasks, and discussion (CfEM, 2023).

The maths mastery approach was originally developed in the UK for teaching in primary and secondary schools. CfEM updated the approach for GCSE resit and FSQ Levels 1 and 2 teaching for 16-19-year-old learners in Further Education (FE) settings.

Previous experimental and quasi-experimental trials of mastery interventions have found promising results. An assessment of evidence from previous studies, including quasi-experimental evaluations, meta-analyses and observational studies, pointed to mastery practices that showed promise for improving maths attainment (Boylan et al., 2018). Further, in 2015 the Education Endowment Foundation (EEF) funded 2 RCTs of Maths Mastery in England, one for Year 1 pupils (Vignoles et al., 2015) and one for Year 7 pupils (Jerrim et al., 2015). Although the trials showed mixed results, a meta-analysis revealed an overall statistically significant positive impact (Jerrim & Vignoles, 2015). Moreover, the intervention in these 2 EEF-funded studies lacked certain elements of Maths Mastery which have been included in other trials of the intervention, with meta-analysis of these interventions showing higher estimated effect sizes (Guskey and Pigott, 1988; Kulik et al., 1990).

The efficacy of the mastery approach has also been evaluated in a randomised controlled setting for older participants. Wake et al. (2023) conducted a trial implementing a teaching-for-mastery approach that aimed to enhance post-16 learners' attainment in Level 2 mathematics. Although the trial faced challenges related to Covid-19 disruptions, the trial found small effects in the treatment groups (two levels of treatment were explored: full and partial), with an estimated one-month learning gain for the group receiving the full intervention. No studies were found testing maths mastery approaches in adult (19+) education settings.

1.1.3. Rationale

Despite promising evidence of maths mastery approaches in school and post-16 learners, there is a lack of evidence from adult education settings. Departing from the standard curriculum delivery approach could be particularly beneficial for adult learners who have often had negative previous experiences in school. This trial therefore sought to test whether AMA, a new intervention developed by ETF, which draws on and further adapts the mastery approach, may be suitable for adult (19+) FSQ Level 1 maths learners.

There is ongoing and longstanding interest in improving levels of adult numeracy in the UK population. While Functional Skills Level 1 courses are relatively popular, failure in maths, leading to maths anxiety, is not uncommon and this undermines confidence to take part in learning to rectify this. A mastery approach for FSQ Level 1 aimed to provide learners with a deeper understanding of maths and provide them with a greater range of approaches and strategies.

1.2. Intervention description: theory of change

The AMA intervention aimed to improve adult learners' confidence and achievement in Level 1 maths by adopting a mastery approach, adapted to adult numeracy settings.

The Theory of Change envisaged that financial resources and mastery lesson/training resources (inputs) would support recruitment of learners and tutors, and training of tutors and Continued Professional Development (CPD) (activities). It was expected that these activities would enable delivery of 15 maths lessons centred on maths mastery approaches, encouraging learners to engage in critical thinking and to understand maths at a deeper level (outputs).

It was hoped these lessons would enable learners to understand maths concepts more deeply and to understand how different maths topics connect to one another. It was also hoped that interactive teaching methods, questioning, short tasks and discussion would promote learner engagement and motivation, and that an inclusive approach to teaching would encourage self-belief and help learners overcome anxiety stemming from prior negative experiences (change mechanisms). As a result, it was hoped that tutors would change their attitudes towards learner capabilities and promote a more positive mindset, enabling learners to make connections between maths topics, becoming increasingly willing to learn and adopt a more positive attitude to maths, and to have a deeper understanding of maths concepts (short-term outcomes). It was hoped these changes in attitude and knowledge would help learners to strengthen their maths ability, achieve FSQ level 1 maths and improve their confidence and self-belief in maths (medium-term outcomes).

Ultimately, the intervention aimed to contribute to:

- Improved functional numeracy across the adult population in England;
- Greater retention and achievement in FSQ pathways, along with long-term improvements in employability, confidence and independence, contributing to a better skilled workforce;
- Wider social benefits such as individuals being more able to support with children's learning and manage personal finances; and
- Strengthened adult numeracy provision across the system

This Theory of Change is summarised below in Table 1.

Table 1: Theory of Change

Situation	There is longstanding policy interest in improving levels of adult numeracy in the UK population. Levels of participation and achievement in government-funded courses falling in the decade prior to Multiply	Aims	AMA aims to improve learner attainment and confidence in maths by focusing on teaching a deeper understanding of maths than in the standard curriculum.
------------------	--	-------------	---

Inputs and activities	Outputs	Change mechanisms	Outcomes	Impacts
<p>Inputs</p> <ul style="list-style-type: none"> • Financial resource • Maths mastery lesson resource <p>Activities</p> <ul style="list-style-type: none"> • Recruitment of learners and tutors • Tutor training, ongoing support and CPD 	<ul style="list-style-type: none"> • No. tutors recruited • No. training sessions delivered • No. learners recruited • Tutors deliver 15 lessons, adopting key teaching principles of the mastery approach including critical thinking and providing a deeper understanding of maths 	<ul style="list-style-type: none"> • Exposure to a sequenced curriculum and encouragement to think critically help learners understand topics deeply and how they connect • Interactive teaching, questioning, short tasks and discussion builds learner engagement and motivation • Novel and inclusive approach to maths teaching helps learners' self-belief to overcome anxiety from prior negative experiences 	<p>Short term</p> <ul style="list-style-type: none"> • Tutors change their mindset towards learner's capabilities to learn maths • Learners make connections between topics • Learners are increasingly willing to learn and improve attitude to maths • Learners have deeper and improved understanding of maths <p>Medium term</p> <ul style="list-style-type: none"> • Learners achieve FSQ maths level 1 • Learners have improved confidence in maths and self-efficacy compared to baseline • Learners have better ability in maths compared to baseline 	<ul style="list-style-type: none"> • Improved functional numeracy across adult population in England, with more attainment at level 1 FSQ • Better skilled workforce, with increased learner employment and progression opportunities • Social/economic benefits of individuals being able to support children's learning/manage personal finances • Strengthened adult numeracy provision across the system

<p>Evidence assessment</p>	<p>The intervention was informed by evidence of positive effects of mastery approaches among school-aged learners and post-16 learners. No research was found on the effects of mastery approaches for adult maths learners in England undertaking FSQ Level 1 maths.</p>		
<p>Assumptions</p>	<ul style="list-style-type: none"> • Tutors are given enough time to fully engage with the intervention both at initial training and during ongoing weekly/fortnightly online sessions. 	<p>Possible unintended consequences</p>	<ul style="list-style-type: none"> • The intervention may be less suitable for some groups of learners or online delivery, potentially disadvantaging some who would benefit from standard teaching approaches. • Some learners may struggle to move on to further learning which is based on traditional methods.

1.3. Intervention description⁴

Name

Adapted Mastery Approach for Functional Skills Qualification Level 1

Why: Rationale, theory and/or goal of essential elements of the intervention

The maths mastery approach developed by the CfEM programme asserts that all individuals can enjoy and excel in mathematics. The curriculum is designed to provide a coherent and detailed sequence of essential content, ensuring sustained progression over time. Lessons are intricately linked to prior learning, carefully sequencing steps to build a secure understanding, and various instructional methods are employed, such as whole-class interactive teaching, questioning, short tasks, and discussion (CfEM, 2023). Departing from the standard curriculum delivery approach could be particularly beneficial for adult learners who have often had negative previous experiences in school.

See sections 1.1 and 1.2 for more detail.

Who: Recipients of the intervention

The intervention was designed to be delivered in adult (aged 19+) FSQ Level 1 maths classes by tutors teaching these courses who are trained in AMA. While classes could be mixed ability, only learners working towards FSQ Level 1 were eligible for the trial.

What: Physical or informational materials used in the intervention

Tutors received lesson plans, slides and handouts for each of the 15 topic areas of the AMA course. The materials for the intervention focus on the following 5 key principles for mastery of maths, developed by CfEM (2023):

- Teaching that allows learners to develop and understanding of mathematical structure.
- Valuing and building on learners' prior learning.
- Prioritising curriculum coherence and connections.

⁴ This intervention description uses the Education Endowment Foundation (EEF) adapted version of the Template for Intervention Description and Replication (TIDieR). Initially used for health trials, this template is increasingly used in other forms of research for replicability.

- Developing both fluency and understanding of key ideas.
- Developing a culture in which everyone believes everyone can succeed.

Following completion of initial training, tutors were given access to all resources via a Padlet link provided.

What: Procedures, activities and/or processes used in the intervention

The intervention comprised 15 sessions of 1.5 hours over 15 weeks. Each session drew on the 5 principles of maths mastery, as detailed in the previous section on materials.

Who: Intervention providers/implementers

The AMA intervention was developed by ETF, which is a UK sector-led body that aims to support and improve the quality of teaching, leadership, and workforce development across further education and skills training. It was developed under ETF's CfEM programme.

DfE appointed a Managed Service Supplier (MSS), Etio, to lead on recruitment and management of providers to the Adult Numeracy Trials, including AMA.

The intervention was delivered by FSQ Level 1 maths tutors in providers who had been assigned to the AMA treatment arm of the trial. Treatment tutors responsible for delivering the course were required to attend initial training and ongoing study sessions delivered by ETF. The training was delivered across 4 sessions of 2.5-hours each (10 hours total). Trainers also hosted weekly 2-hour lesson study sessions throughout intervention delivery to offer ongoing support to tutors, who were required to attend at least 80% of these.

How: Mode of delivery

The intervention was designed primarily for face-to-face delivery, but providers that delivered courses online were also able to participate in the intervention and the trial. Based on responses to the endline learner survey, 69% of learners said that their course was delivered in person only, 20% said their course was delivered online only, and the remainder (11%) said that their course was delivered as a mix of in person and online delivery.

Where: Location of the intervention

Providers were recruited from across England. The setting for AMA was primarily adult education and FE colleges, as well as some online learning environments.

When and how much: duration and dosage of the intervention

AMA was designed to be delivered for 1.5 hours per week for 15 weeks. The maximum dosage was therefore 22.5 hours. For FSQ Level 1 maths there are up to 55 Guided Learning Hours (GLH).

The frequency of classes and intensity of dosage varied according to providers' schedules. In some cases, learners received a lower weekly dose of the intervention over 30 weeks. The intervention commenced at the beginning of the 2024/25 academic year and continued until either Christmas 2024 or summer 2025, depending on the provider's timetable. The total number of hours to be delivered was 22.5 irrespective of the calendar time taken to deliver the course. Learners were expected to receive at least 80% of the AMA lessons, or 18 hours of teaching time.

Some participants began classes after the end of the 15-week intervention cohort in January 2025. In these cases, ongoing support was not provided by ETF for the new cohort, but they were still available for queries or when providers needed support. Data was still collected.

Tailoring: Adaption of the intervention

The delivery model could be customised to align with the different types of FSQ Level 1 delivery models adopted by participating providers. As mentioned, providers had the choice to deliver the 15-week, 1.5-hour per week schedule or spread the 22.5 lesson hours over 30 weeks depending on their scheduled curriculum.

However, customisation of the content itself was not part of the delivery plan. Tutors were required to adhere to the content as established by the product developer. Adjustments to the content were to be kept to a minimum with the aim of delivering the entire content without exception or alteration.

Modifications: Changes to the planned intervention

There were no modifications to the planned intervention.

Strategies to maximise effective implementation

Fidelity was assessed through the implementation and process evaluation (IPE), using the quantitative surveys and qualitative evidence from tutor and learner interviews.

Evidence of implementation variability

The intervention was delivered with good overall fidelity; more information can be found in section 4.3.

1.4. Evaluation overview

This pilot RCT gathered evidence on the quality, relevance, suitability, and effectiveness of the content of the AMA course in adult education settings and the processes involved with implementing it, as well as any adaptations that might be required to support future delivery or wider roll out. It tested the effectiveness and feasibility of adopting maths mastery approaches to teaching within adult education settings, specifically FSQ Level 1 maths courses.

Feeley and colleagues (2009) proposed that feasibility studies (in this case, a pilot study) gauge both the capability to effectively deliver the intervention and to carry out the evaluation design. To date, relatively few RCTs have been conducted in the UK adult education sector, resulting in limited robust evidence on the effectiveness of interventions aimed at improving adult skills. The learning gathered through the pilot was intended to provide additional value in respect of informing the design and delivery of full scale RCTs within the sector in the future.

The focus of the impact evaluation was on testing whether the intervention achieved its intended outcomes. It used an experimental design to assess the causal effect of using an adapted mastery approach to teach maths in adult education setting on the outcomes of interest. As part of a 3-arm trial, the intervention was compared to another intervention (Contextualised Approach) and also to BAU, i.e. the control group, who received FSQ Level 1 maths courses without either intervention.

For the IPE, evidence was gathered on the processes involved in the delivery of the intervention including quality, relevance and suitability, as well as the experiences of adult education organisations and adult learners in participating in a pilot RCT.

DfE commissioned the Policy Institute at King's College London to deliver this evaluation.

1.5. Research questions

1.5.1. Impact evaluation

Primary research question:

- What is the average difference in **FSQ Level 1 maths attainment**, measured by pass rate (binary outcome of pass/fail) of adult learners in providers that deliver AMA as part of their FSQ Level 1 maths course, in comparison to adult learners who do not receive AMA as part of their FSQ Level 1 maths course?

Secondary research questions:⁵

- What is the average difference in **attendance**, measured using records by providers, of adult learners in providers that deliver AMA as part of their FSQ Level 1 maths course, in comparison to adult learners who do not receive AMA as part of their FSQ Level 1 maths course?
- What is the average difference in **changes in confidence** in maths, measured by baseline and endline surveys, of adult learners in providers that deliver AMA as part of their FSQ Level 1 maths course, in comparison to adult learners who do not receive AMA as part of their FSQ Level 1 maths course?
- What is the difference in **maths skills**, measured using an endline maths test, of adult learners in providers that deliver AMA as part of their FSQ Level 1 maths course, in comparison to adult learners who do not receive AMA as part of their FSQ Level 1 maths course?

1.5.2. Implementation and process evaluation

Main IPE research questions:⁶

1. Has the intervention been delivered with fidelity (that is, in line with the intervention guidance)?
2. Is the pre-delivery CPD training delivered as intended, and is it consistently delivered?
3. To what extent, if at all, do BAU FSQ Level 1 courses typically include any content relating to AMA?
4. Is there any evidence of contamination of BAU FSQ Level 1 courses, or contamination of one of the treatment arms with the other treatment arm? If so, what are the causes of this?
5. How did tutors experience delivering the intervention?
6. To what extent do tutors integrate AMA principles, approaches and activities into their lessons? To what extent did they make changes to their lesson plans? Did this change over time?
7. What are tutors' experiences of the training and support provided to deliver the intervention?
8. What are learners' experiences of the intervention?

⁵ Note that the questions on attendance and maths skills were not included in the final analysis. Information about variations from the protocol can be found in section 2.1.8.

⁶ The full set of IPE research questions and corresponding sub-questions are detailed in Appendix 4.

9. What motivates learners to take or re-take the FSQ Level 1 teaching course?
What are they hoping to gain?
10. What outcomes does the intervention have for learners?
11. What enablers and barriers are there to learner engagement and participation in the intervention?
12. What lessons have been learned from future delivery of the intervention or wider rollout?
13. What lessons can be learned from the pilot about delivering RCTs in adult education settings?
14. Were the outcome measures used valid, reliable, acceptable and complete? How / in what ways could they be improved?
15. What were the reasons for any attrition from the pilot trial? How did attrition vary between treatment and control groups and different types of learning organisations / learners?

1.6. Reading the report

This report outlines the evaluation methodology, findings from both the impact evaluation and IPE and provides conclusions and recommendations. A glossary of technical terms can be found in Appendix 1.

Further details about the intervention, costings, data collection tools, code used and additional data tables can be found in Appendices 2-6.

2. Methods

This section describes the methods used for the impact and implementation and process evaluations.

2.1. Impact evaluation

2.1.1. Trial design

Table 2 provides an overview of the trial design as outlined in the Pilot Trial Protocol.⁷

Table 2: Trial design table

Trial element	Description
Trial name	Embedding the use of an Adapted Maths Mastery Approach, or a Contextualised Curriculum Approach, in Functional Skills Level 1 maths courses for adults 19+: a three-armed cluster randomised pilot trial
Project title	Adapted Mastery Approach for FSQ Level 1 in Maths
Developer	Education and Training Foundation (ETF)
Geography	England
Delivery	Adult learning providers
Evaluator	King's College London, Institute for Employment Studies
Principal Investigators	Becci Newton (Lead), Michael Sanders
Trial Protocol authors	King's: Michael Sanders, Susannah Hume, Dimitris Vallis, Gabrielle McGannon IES: Seemanti Ghosh, Becci Newton, Emily Kramers, Susie Bamford
Trial design	3-armed parallel cluster randomised pilot trial with random allocation at provider level First treatment arm: Adapted Mastery Approach intervention Second treatment arm: Contextualised Approach intervention Control arm: FSQ Level 1 maths courses without either intervention

⁷ The pilot trial protocol can be found here: <https://osf.io/vpu39>

Trial element	Description
Trial type	Pilot gateway RCT
Trial population	Adults (19+) completing a Functional Skills Level 1 qualification in maths
Primary outcome	Functional Skills Qualification Level 1 pass rate
Secondary outcomes	Attendance, confidence in maths, maths attainment

The impact evaluation was delivered as a 3-arm pilot RCT with randomisation at the provider level. The primary outcome was attainment (pass/fail) in FSQ Level 1 maths courses. The primary analysis followed an intention-to-treat (ITT) approach using a multilevel logistic model, per the protocol. Given that the primary outcome was a binary variable, the estimated treatment effect is reported as a percentage point (ppt) change in the probability of treated individuals passing the FSQ Level 1 course compared to the control. Pilot RCTs are usually only statistically powered to detect effect sizes that are unrealistically large with some level of certainty. The sample size for this trial was not expected to enable an indication of the impact of the intervention and its effect size.

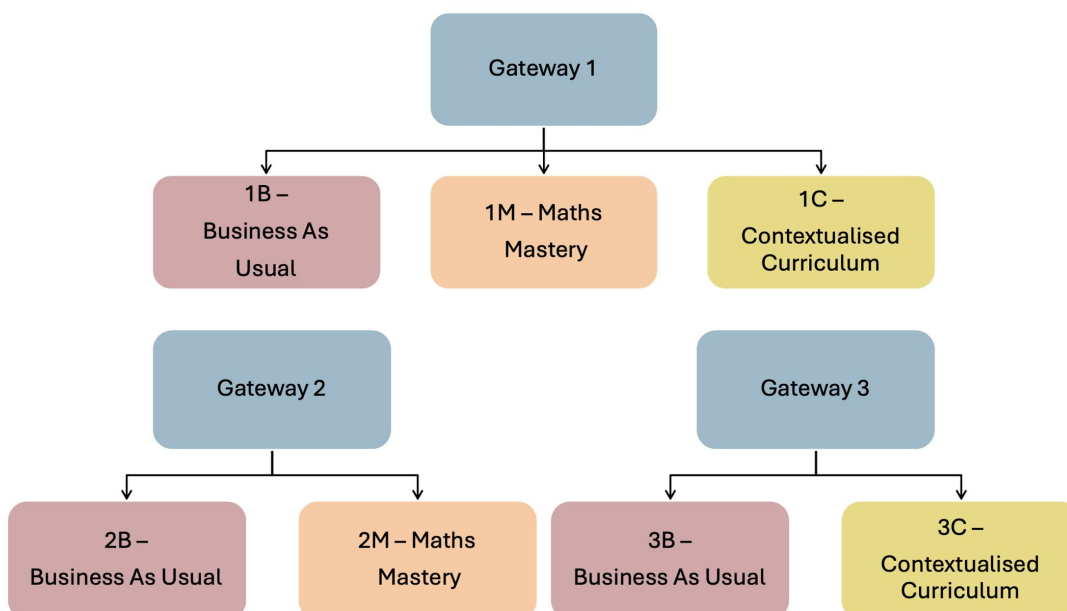
For the secondary outcome (confidence in maths), data collection was conducted at baseline and endline, determined by the length of the curriculum (15 or 30 weeks depending on provider). The other secondary outcomes (attendance and maths skills) were not included in the final analysis due to changes to data collection.

This pilot trial was constructed by merging 2 trials that were previously intended to run individually (Contextualised Approach and AMA). Having a single pooled control group for the 2 interventions maximised the sample size and statistical power given the constraints in recruiting providers. The control group received FSQ Level 1 maths courses without either intervention. One key consideration when merging the trials was that providers might have strong preferences for one intervention over the other, to the extent that they would rather not participate at all than face a 1 in 3 chance of being assigned to their non-preferred intervention. The MSS was therefore able to offer participation in a 2-armed pilot RCT without the intervention the provider did not want to implement (e.g. Contextualised Approach and control; or AMA and control) to providers who would otherwise withdraw from the trial.

This design is called a gateway trial, because providers were first offered the 3-armed trial and were then only offered 1 of the 2-armed trials if they were so opposed to an intervention that they would rather withdraw from the 3-armed trial than risk being allocated to it. This approach was also taken if providers had already been exposed to one of the interventions. In that case, they were moved from the 3-armed trial to the 2-armed trial for the intervention they had not been exposed to.

Providers who chose to progress to the second gateway, and therefore were excluded from being randomised to 1 of the 2 possible treatments, had a lower probability of receiving an intervention (66% chance of receiving an intervention in the 3-armed pilot trial vs. 50% chance of receiving the intervention in the 2-armed pilot trial). The evidence base for both interventions is also largely equivalent, so providers were not gaining any significant benefit by progressing to the second gateway. It was agreed that if a provider found out about another provider’s gateway selection and raised concerns, the MSS would explain that the interventions are considered equal in their likelihood of impact and therefore all providers were initially given the option that gave them the highest chance of being randomly allocated to one of the interventions. The flow of the gateway design is shown in Figure 1.

Figure 1: Gateway trial design



In total, no providers elected to go through gateway 2, while 12 providers chose gateway 3 (the Contextualised Approach option), with the majority choosing this route due to previous exposure to maths mastery interventions.

2.1.2. Recruitment

Recruitment was led by the MSS during summer 2024. Adult learning providers were recruited from among the 480 adult education providers delivering FSQ Level 1 maths courses. Tutors were recruited through the participating providers. They were provided with information about the trial and were able to opt-out if they did not want to participate.

Learners were also recruited through the participating providers. The inclusion criteria were learners aged 19+ studying for FSQ Level 1 maths at a participating adult education provider in the 2024/25 academic year. Learners were provided with information about

the trial and were able to opt-out if they did not want to participate. Learners who opted out of the trial were still able to access the course, but they were not invited to take part in any further evaluation activities, and their data was not included in the trial.

2.1.3. Sample size

This trial was undertaken as a pilot with a small number of participants. It was not sufficiently powered to detect small to moderate effects. The sample size was understood from the outset to be too small to detect the expected impact of the intervention.

A minimum recruitment target of 66 providers (for 22 providers per arm across the 3-arm trial) was set. This meant that, if all providers elected for the second gateway, there would be 30 providers per individual pilot trial after provider-level attrition (estimated to be 30%). The sample of providers recruited and randomised as part of the multi-arm trial was 19 for the AMA arm, 19 for the Contextualised Approach arm, and 19 for control (shared between both arms).

Therefore, for the AMA analysis, the total sample size consisted of 38 providers (19 AMA treatment and 19 control) prior to any post-randomisation dropouts. However, 13 providers withdrew from the trial, leading to a final analysis sample size of 25 providers across both AMA treatment (13) and control (12) groups. The total learner sample size included in the analysis for AMA was 914,⁸ with 444 learners in control and 470 learners in treatment across the 25 providers.

Attrition was substantial across both outcomes. For the primary outcome (pass rate for FSQ Level 1), approximately one third of providers were lost to follow-up (34%). This was due to dropouts as well as difficulties encountered in matching records and identifying outcomes for learners, with similar attrition rates in the treatment and control groups.

Attrition was higher for the secondary outcomes, affecting nearly half of participants (45%), again with comparable rates across groups. Overall, attrition did not differ substantially by treatment status.

Variation of cluster size by provider was high, with an average of 37 learners per provider and a standard deviation of 60. The smallest cluster size had 2 learners and the largest had 276, suggesting large variation in class sizes by provider. The 3 largest providers contributed 55% of the total learner sample used in the analysis. This variation, along with the effects of attrition, potentially resulted in a substantial reduction in statistical power, limiting the levels of inference.

Given the post-attrition sample of 25 providers, with an average cluster size of 37 learners, baseline-endline correlation of 0.2 and an intracluster correlation (ICC) of 0.19,

⁸ During analysis, the sample size was reduced as a result of missingness stemming from other covariates.

this pilot trial was powered to detect a difference of 14 ppt from a baseline of 51%, or a Cohen's h of 0.51.

2.1.4. Randomisation

The randomisation was carried out across all 3 arms and results were shared with the Institute for Employment Studies (IES) who led the Contextualised Approach arm of the trial. Adult education providers were randomised to treatment (Contextualised Approach or AMA) and control groups using a randomisation algorithm in Stata that aimed to create 3 equally sized arms (see Appendix 5 for the randomisation code). Randomisation was stratified by the region in which the provider was operating, across all 9 regions of England. Balance on baseline characteristics between treatment and control was then assessed using the inclusion of the stratification variable (region) to ensure randomisation was successful.

Where a provider opted to take part in one of the gateway trials, the same stratified randomisation procedure was used, but with the outcome limited to 2 arms (the relevant intervention and control) rather than all 3 arms. During recruitment, the second gateway was only used where providers had previously been exposed to a maths mastery intervention or preferred to be randomised to the Contextualised Approach or control arms only. Therefore, Gateway 2 was never activated. The analysis for this pilot trial only used controls from Gateway 1 and excluded providers randomised to Gateway 3, per the protocol.

Randomisation and tutor training were conducted in separate batches of providers due to the rolling nature of recruitment. Randomisation was conducted in 6 batches of recruited providers. Once a provider was allocated to the AMA treatment arm, the relevant tutors at that provider were invited to AMA training sessions. Tutors at providers assigned to the control condition were not invited to any training sessions. All learners enrolled on a FSQ Level 1 maths course at a participating provider were assigned to the same condition as the provider.

2.1.5. Outcome measures

Primary outcome measure

The primary outcome measure was the pass rate for FSQ Level 1 maths. The evaluation used the Outgrade variable in the ILR to measure pass/fail for learners included in the trial. Learners were able to take the exam multiple times over the academic year. A cut-off date was set based on the delivery model chosen by each provider (15 weeks or 30 weeks) and the overall pass rate after that cut-off (within the 2024/25 academic year) was used as the primary outcome. For providers that ran multiple cohorts within the trial period, all cohorts were included in the analysis. The study did not expect learners to be influenced by treatment assignment when deciding whether to enrol on a class as they would not be aware of the assignment until after enrolment.

However, to account for the fact that treatment assignment may have affected a learner's likelihood of continuing with the course, any participant who enrolled but dropped out or did not complete their course was counted as having failed (score imputed as 0). In total, of the 914 learners in the matched records, 339 did not have a pass/fail outcome available within the ILR, 47 in control and 292 in treatment. This imbalance was primarily driven by a provider randomised to treatment that had a large number of participants whose outcomes were missing, representing 73% of all missing outcome cases.

Secondary outcome measure

The only secondary outcome included in the final analysis was confidence in maths. The other 2 secondary outcomes in the protocol (maths skills and attendance) were not included because the endline maths skills test was not administered and attendance data was not collected from control providers (see section 2.1.8 for information on changes from the protocol). Maths confidence was collected via a survey measure at baseline and endline. The main confidence measure was a Likert question asking about learners' overall level of confidence in maths, ranging from 1 (Not at all confident) to 5 (Very confident). This measure was used across the Adult Numeracy Trials.

Confidence in maths: Overall, how confident do you feel working with numbers in everyday life?

A composite measure of confidence in maths was also included in the analysis, taking the average score of all non-missing items from the following 6 Likert-type questions with possible responses ranging from 1 (Not at all confident) to 5 (Very confident). This composite measure was in addition to the analysis specified in the protocol, which only included the single confidence in maths question. The detailed questions are included in Appendix 4.

Confidence in maths (composite): And how confident do you feel about doing the following things in everyday life?

- a) Checking your change is right when you have bought something.
- b) Working out the best deals when shopping.
- c) Helping children with maths homework or talking about maths / numbers with children (if respondent is a parent).
- d) Understanding interest rates on bank statements.
- e) Keeping track of your bank account balance.
- f) Working with numbers as part of a job (if respondent is working).

2.1.6. Data collection

Primary outcome

Data for the primary outcome was collected via the ILR.

Secondary outcome

Learners in both treatment and control groups were invited to participate in a baseline and endline survey. These surveys, administered by Ipsos, were conducted both online (distributed to learners through providers) and as telephone surveys (using contact details shared by providers to contact learners directly) to maximise response rates. The start and end dates were aligned to course start and end dates, which were staggered across providers during the trial. The baseline survey was open between 20 September 2024 and 26 March 2025, while the endline survey was administered in waves from 27 November 2024 to 11 September 2025.

The total number of responses are shown in Table 3. The estimated response rates at baseline were similar between treatment and control learners, at 23% and 26% respectively⁹. At endline, however, there was a much higher response rate from control learners (31%) than treatment learners (13%), and the response rate from control learners was higher at endline than baseline. It is unclear why these differences occurred.

Table 3: Number of responses and response rates for learner surveys

Survey	Treatment/ control	Number of responses (n)	Response rates	Total learners
Baseline	Treatment	107	23%	470
Baseline	Control	114	26%	444
Endline	Treatment	62	13%	470
Endline	Control	137	31%	444

Source: baseline and endline learner surveys.

⁹ Response rates usually reflect the number of learners who completed a survey divided by the number of learners who were invited to take part in the survey x 100. As the baseline surveys for AMA were distributed by providers via an open link, we cannot say with certainty the total number of learners who received the invitation. As such, the response rate is an estimate based on dividing the number of responses by the number of learners who were in the trial at the time of randomisation x100. The true response rate could be higher or lower.

2.1.7. Statistical analysis

Primary analysis

The primary analysis followed an intention-to-treat (ITT) approach. Analysis of the primary outcome was conducted using a multilevel logistic model of the form:

$$Y_{ij} = \Lambda(\beta_0 + \beta_1 T_j^M + \beta_2 X_{ij} + \beta_3 S_j + u_j)$$

Where cluster variance was defined as:

$$u_j \sim N(0, \sigma_u^2)$$

Where $\Lambda(\cdot)$ is a logistic link function, Y_{ij} is the binary outcome of pass/fail for participant i with provider j , T_j^M is the treatment indicator for AMA, X_{ij} is a vector of relevant individual-level characteristics, S_j represents stratum in which cluster j belongs represented by dummy indicators, and u_j the provider random intercept.

The vector of relevant individual-level characteristics X_{ij} includes the following covariates:

- A binary variable of whether the individual i is Female=1 or Male=0.
- A binary variable of whether the individual is of a white ethnic background=1 or not of a white ethnic background=0.
- A continuous variable for age.
- A binary variable of whether the individual is of ill health or disability=1 or not of ill health or disability=0.
- A binary variable of whether the individual lives in an area with a deprivation score in the 20% most deprived=1 or not=0.
- A binary variable of whether the individual is employed=1 or not employed=0.

Secondary analysis

Analysis of the secondary outcome of confidence in maths followed a linear autoregressive (AR(1)) model specification of the form:

$$Y_{ij} = \beta_0 + \beta_1 T_j^M + \beta_2 X_{ij} + \beta_3 S_j + \beta_4 Y_{t-1,ij} + u_j + e_{ij}$$

Where the cluster and residual variance terms are defined as:

$$u_j \sim N(0, \sigma_u^2)$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

Where Y_{ij} is the continuous outcome of maths confidence for participant i with provider j , T_j^M is the treatment indicator for AMA, X_{ij} is a vector of relevant individual-level characteristics using the same pre-specified covariates as in the primary analysis model, S_j represents stratum in which cluster j belongs represented by dummy indicators, $Y_{t-1,ij}$ is the outcome at baseline, and u_j, e_{ij} the provider-level and individual-level residuals.

Compliance analysis

Compliance was defined as a learner attending at least 80% of their scheduled FSQ Level 1 maths classes. Learners in the treatment group may not have received all 22.5 hours of the intervention for various reasons, but for the compliance analysis, a learner was considered compliant if they attended 80% of classes (12 out of 15) that were taught by a tutor trained in the intervention as defined in the registered trial protocol.

To model compliance, a binary variable was generated: 1 if the individual met the compliance criteria and 0 otherwise. The analysis initially examined the level of non-compliance through descriptive statistics. The protocol stated that if over 80% of learners were deemed compliant, the study would not conduct further compliance analysis. However, compliance was found to be below 80%, so the Complier Average Causal Effect (CACE) was estimated using an instrumental variable (IV) approach with two-stage least squares (Imbens and Angrist, 1994). Given the smaller sample size of this pilot trial, the CACE estimate should be treated as exploratory. Furthermore, this type of IV approach generally leads to higher standard errors and therefore lower precision.

Missing data analysis

The analysis used administrative data from the ILR for the primary outcome and covariates, making it less susceptible to the impact of missing data than analyses relying on survey data, and the likelihood of extensive missing data was lower.

Where missingness exceeded 5%, patterns in demographic characteristics and auxiliary variables were examined to determine whether the data was Missing Completely at Random (MCAR) or potentially Missing at Random (MAR). Binary indicators for missing/non-missing (1/0) were generated and logistic regression models were used to examine whether observed covariates could predict missingness. Where appropriate, Multiple Imputation by Chained Equations (MICE) was performed using covariates that supported the MAR assumption (White et al., 2011).

Sub-group analysis

Exploratory subgroup analyses were conducted on the primary and secondary outcomes to assess whether the impact of the intervention showed any evidence of heterogeneity in effect between the following subgroups:

- Sex: Baseline ILR data using the binary variable SEX.

- Learning difficulty/disability or health issue: Baseline ILR data using the categorical variable LLDDHEALTHPROB.
- Ethnicity: Baseline ILR data using the categorical variable ETHNICITY (White vs other ethnic backgrounds).

Estimation of effect size

As the primary outcome was the pass rate for FSQ Level 1 maths, the study used historical data to shape its assumptions. The overall pass rate for FSQ Level 1 maths was reported as 53% for the 2022/23 academic year and 48% for the 2021/22 academic year (NCFE, 2023).

Cohen's h was estimated using the following formula:

$$h = 2 \arcsin(\sqrt{p_1}) - 2 \arcsin(\sqrt{p_0})$$

Following estimation of the logistic regression, the respective proportions were then calculated as the probability of FSQ achievement, estimated from the main model as the average predicted probability if everyone were untreated:

$$p_0 = \Lambda(\beta_0 + \beta_2 X_{ij} + \beta_3 S_j + u_j)$$

And the average predicted probability if everyone were treated:

$$p_1 = \Lambda(\beta_0 + \beta_1 T_j^M + \beta_2 X_{ij} + \beta_3 S_j + u_j).$$

With the upper and lower tails of Cohen's h confidence interval estimated as:

$$h \pm 1.96 \sqrt{\frac{2}{n}}$$

For the effect size of the secondary outcome (maths confidence), the effect size was estimated using Hedge's g formula:

$$ES = \frac{\beta_1}{\sqrt{\sigma_u^2 + \sigma_e^2}}$$

Where β_1 is the treatment effect estimate and σ_u^2, σ_e^2 the estimated cluster and residual variances, respectively.

Additional analysis and robustness checks

In addition to the ITT, analysis was conducted to gauge the effects on the estimated impact as a result of the exclusion of those whose outcome was imputed as a fail (0) due to a lack of examination. This analysis was not part of the protocol as it was purely data-driven and observed during analysis and therefore required an ad-hoc approach.

2.1.8. Variations from protocol

The following changes were made to the final analysis compared to the trial protocol.

Of the 3 secondary outcomes specified in the protocol, 2 were not included in the analysis due to lack of data.

- Maths skills: the trial protocol included an endline test of maths skills for control and treatment learners. However, this was not delivered as planned. The maths assessments were originally intended to be administered at baseline and endline to measure progress over time. However, due to timing constraints (including course duration) and low uptake of the optional assessments, this was not possible. The maths assessments were instead administered to examine how the tests could work in practice (“testing the test”) and to capture learning on the administration and validity of the tests as a research tool.
- Attendance: this analysis required attendance data from both treatment and control groups. However, as no attendance data was collected from the control group, it was not possible to proceed with the analysis.

Additionally, a “composite” maths confidence score was added to the analysis, compiled from a range of questions in the learner surveys that related to maths confidence (full questions can be found in Appendix 4). This was included as a secondary measure for the confidence outcome.

2.1.9. Limitations

One of the main limitations of the pilot trial was the sample size. The final effective sample size used in the analysis was lower than the target of 15 providers per arm (30 per trial), which weakened the precision of standard errors corrected for clustering.

Another significant limitation was incomplete outcome data. Some participants who attended courses and were part of the trial did not have “Functional Skills Qualifications in mathematics” recorded as their aim in the ILR. This meant it was not possible to identify their relevant Pass/Fail outcome within the “Outgrade” variable, so these had to be imputed as 0 (fail) for 339 participants. This almost certainly attenuated the estimated effect of the intervention by biasing it downward, even under CACE.

Finally, since matching was not exact throughout, any imprecision in matching may also have contributed to additional noise in the data, reducing precision. However, this effect is likely to be minimal as the matches used met the required minimal threshold.

2.2. Implementation and process evaluation

The IPE provided insights into the causal pathways for AMA for FSQ Level 1. The IPE aimed to:

- Provide evidence on how AMA was delivered and integrated into the existing FSQ Level 1 maths curriculum, the extent to which it was delivered consistently across providers and tutors, in the manner specified, and to what extent and how it differed from usual practice.
- Explore tutor and learner experiences including views on what worked well/less well in delivering the intervention, to identify lessons for potential future rollout.
- Explore perceived outcomes for learners and tutors that were anticipated to result from the intervention, but which were not a focus of the pilot RCT.
- Investigate how outcomes and impacts emerged, and were perceived among tutors and learners, including unanticipated outcomes and any difference made to their experience and practice.
- Gather insights into the feasibility and acceptability of delivering RCTs in adult education settings, including enablers and barriers to participation, levels of attrition, compliance with data collection tools and approaches and experiences of tutors and learners in both treatment and control groups.

The IPE used a mixed-methods design, collecting data from learners, tutors, and providers. This included:

- Baseline and endline surveys of learners in the treatment and control groups.
- Qualitative interviews with learners in the treatment and control groups.
- Qualitative interviews with tutors in the treatment and control groups.
- Qualitative interviews with representatives of participating providers in both the treatment and control groups.
- Management information covering tutors' attendance at training sessions and learners' attendance at course sessions.

The interviewers used specific topic guides for each group (Appendix 4).

2.2.1. Data collection

Surveys

The learner surveys discussed in the Secondary outcome section also included questions specifically related to the IPE. These covered learners' motivations for taking the course, overall satisfaction with the course, perceived outcomes related to attitudes, willingness to attempt maths, understanding of mathematical structures, awareness and understanding of AMA principles, recall of AMA activities during the course, and suggestions for improvements to the course.

Responses at baseline and endline were not necessarily from the same learners, which means comparisons between baseline and endline should be treated with caution.

Comparisons between the treatment and control learners should also be interpreted carefully, given the relatively low response rates (Table 3).

Interviews

Table 4 details the samples of learners, tutors and providers who participated in qualitative interviews for the IPE.

Table 4: Overview of IPE interviews

Interview participants	Treatment/control	Number of interviews
Learners	Treatment	30
Learners	Control	10
Tutors	Treatment	33
Tutors	Control	10
Providers	Treatment	17
Providers	Control	10

Source: internal interview tracking data (King's and IES).

Base: N/A

Note: some participants had 2 roles relevant to the evaluation. For instance, some had both oversight of delivery and implementation (providers) while also teaching courses (tutors). In these cases (5 participants), to reduce interviewee burden, both interview guides were covered in one sitting that lasted slightly longer. The table above has double-counted these interviews. Therefore, in total, 105 unique interviews were conducted, while the sum in the table above is 110.

Note: all control provider interviews were conducted by IES; the analysis of those interviews was completed by King's for this report.

It should be noted that initial interest in the learner interviews was low. This meant that, while the study reached the target number of learner interviews, it was not always possible to select learners based on target demographic profiles. Therefore, participants were not always balanced on key demographic variables. For instance, for the treatment group:

- Gender: 24 female (80%) and 6 male learners (20%) were interviewed. The aim was for a much more balanced split by gender, but the evaluation struggled to recruit male learners. The numbers were, however, similar to the proportion of learners in the trial. In the sample frame for qualitative recruitment, 160 (66%) were female, while 78 (32%) were male, while in the overall trial, 64% of learners were female and 36% were male.

- Age: the following number of learners were interviewed across different age groups: aged 18-34 (7 learners, 22%), 35-44 (8 learners, 27%), 45+ (15 learners, 50%). This was a satisfactory diversity based on the sample frame for qualitative recruitment: 18-34 (33%), 35-44 (34%), 45+ (33%). In the overall trial, learners were spread across the following age groups: 18-34 (34%), 35-44 (30%), 45+ (36%).
- Providers: the treatment learners came from 8 different providers, with an average of 4 learners from each provider. No learners were interviewed from the remaining 5 providers. Two providers were excluded due to being part of another Adult Numeracy Trial (Embedded Maths) in order to minimise respondent burden, while all eligible learners were invited from the remaining 3 providers but did not respond to the invite.

Learner interviews: treatment and control

Interviews were conducted with 40 learners between February and July 2025. The first wave of the learner survey included a question asking learners whether they were willing to be contacted for a qualitative interview. The sampling frame for learner interviews consisted of those learners who had agreed to be recontacted when asked. It also included learners who did not opt-out of taking part in qualitative research when they joined the trial. This gave us a total of 243 learners across treatment and control groups. Purposive sampling was used to ensure evidence was gathered on the experiences of learners with a range of characteristics and experiences that might affect their experience of the programme. The aim was to recruit diverse participants across genders, ages, and colleges, based on the data available to us at the time of recruitment.

Interviews with learners explored:

- Treatment: motivation to take the course, experience of the course, awareness and understanding of AMA principles, perceived outcomes, enablers and barriers to participation.
- Control: motivation to take the course, experience of the course, perceived outcomes, enablers and barriers to participation.

All learner interviews took place remotely, either via telephone or video call. Each learner interview lasted around 30-45 minutes, and learners who participated in an interview received a £30 online shopping voucher in recognition of their time.

Tutor interviews: treatment and control

Interviews were conducted with 38 tutors between November 2024 and August 2025. Tutors were interviewed in two waves: 15 treatment tutors were interviewed in the first wave (November to December 2024) and 18 treatment tutors participated in the second wave (January to May 2025), with 10 of these tutors having also participated in the first wave interviews. The first wave of interviews took place early in course delivery and the

second wave took place shortly after the end of intervention delivery. Interviews with control tutors took place once, at endline, with 10 control tutors participating in interviews.

Interviews with tutors explored:

- Treatment (wave 1): awareness and experience of AMA principles, view of difference to normal teaching practices, experience of training and support.
- Treatment (wave 2): experience of delivering lessons, fidelity in delivery, enablers and barriers to delivery, learner experiences, perceived outcomes on learners and tutors themselves, experience of taking part in a trial, lessons for future delivery.
- Control: awareness and experience of AMA principles, normal teaching practices, learner experiences, perceived outcomes on learners, experience of delivering lessons, experience of taking part in a trial.

Due to a lower number of course providers than anticipated, the sampling frame was relatively small compared to the target number of interviews. This meant, in practice, that all tutors who agreed to take part were interviewed.

Provider interviews: treatment and control

Interviews were also conducted with senior representatives from adult education providers participating in the trial, from both the treatment and control groups.

Interviews with providers (in both the treatment and control groups) explored: experience of taking part in a trial, experience of trial communications and delivery, experience of training and delivery, perceived outcomes, view of the trial.

As with the tutor interviews, due to the small sampling frame in relation to the target number of interviews, all providers who agreed to take part were interviewed.

Management information

Tutor data was collected via the Ipsos Data Portal using a pre-designed Excel template. Treatment and control group tutors were asked to share their contact details for interview purposes. Treatment tutors were also asked to record when, by whom, and how much of the treatment sessions were delivered to learners.

The following monitoring information was collected and analysed:

- Data from 16 tutors in the treatment group self-reporting whether they had delivered each of the 15 sessions, for each of their cohorts, and the dates of delivery for each session. One tutor reported on 3 cohorts; 4 tutors reported on 2 cohorts, and the remaining tutors all reported on 1 cohort.
- Data from providers about attendance of all treatment learners to each of the 15 different sessions.

2.2.2. Analysis

Analysis of interviews

Interviews were recorded and transcribed by a professional transcription company. All transcripts were then entered into NVivo 14 for data management and analysis. A thematic framework matrix was developed, and data was summarised into this framework. This approach allowed data to be organised under descriptive themes, while retaining the ability to view any individual's journey. Once the qualitative data had been managed, it was analysed descriptively using a process of detection, categorisation and classification (Ritchie and Lewis, 2003). Researchers then conducted explanatory analysis to look for linkages, associations and possible explanatory concepts to provide a rich understanding of how the intervention was delivered, and how different participants experienced it.

Analysis of survey data

Ipsos matched the survey responses to ILR data and learner data collected from providers via the Ipsos Data Portal. All differences between treatment and control group survey responses were tested at the 95% confidence level, with the text explicitly noting whether differences were statistically significant or not.

Survey weights were not applied for several methodological reasons. Firstly, the experimental design with random assignment provided the primary framework for causal inference for the primary and secondary outcomes measures, with randomisation addressing the selection bias that weights typically correct for. Secondly, balance checks showed that survey respondents were broadly representative of the full trial population, despite some demographic differences.¹⁰ Finally, because this was a pilot study with a small number of participants, making statistical adjustments to the data could have made the results less reliable without making them noticeably more accurate. The survey results presented in the IPE findings should therefore be considered indicative rather than definitive.

Analysis of management monitoring information

The attendance data was analysed descriptively using R.

Data triangulation

To triangulate findings, each strand of the data was first analysed separately to ensure that the data was viewed without expectation, then discussing findings as a team to

¹⁰ Statistical comparisons (t-tests and proportion tests) were conducted to examine whether survey respondents differed from the full trial population on key demographic characteristics: age, gender, ethnicity and disability. Whilst most characteristics showed reasonable balance, the tests identified statistically significant differences for disability and ethnicity at endline.

identify convergence and divergence between methodologies. In particular, interview data was re-analysed and interpreted based on the analysis of the survey, aiming to explore how the interview data could explain the survey findings.

2.2.3. Changes from the protocol

The following changes were made to the IPE design compared to the trial protocol.

The endline survey of tutors in the treatment group was removed to reduce burden on tutors, and because it was deemed impractical. The qualitative baseline and endline tutor interviews already captured views from a very large proportion of tutors from the treatment group, so it was agreed with DfE that an additional survey would provide limited added value.

Observations of tutor training sessions were removed. ETF did not feel these were appropriate and were concerned they would affect participation given the small-group format.

2.2.4. Limitations

Survey data

The IPE survey data was subject to non-response, with differential response patterns potentially introducing some bias into the exploratory findings. However, statistical comparisons revealed that survey respondents were broadly representative of the full trial population across available socio-demographic characteristics, despite some differences in disability and ethnicity at endline. Given this broad representativeness and the experimental design of the trial, the decision not to apply weights (as outlined in the methodological approach) was appropriate. However, the IPE survey findings should be considered indicative rather than definitive, providing valuable insights whilst acknowledging potential bias from unobserved characteristics.

Interview data

For tutor and provider interviews, a strong sample was obtained, representing a very large proportion of participants across the trial. However, the number interviewed was below the target numbers in the trial protocol, as the protocol had anticipated a larger number of providers would participate in the trial.

Recruiting learners for interviews proved challenging. Although the target number of interviews was achieved, the sample could have been more diverse. Additionally, the learners interviewed represented a relatively small proportion of the learners in the trial as a whole. The findings therefore may not reflect the views of the wider population of learners. Nevertheless, the qualitative data provides useful insights into the range and diversity of views and experiences. The alignment between learner and tutor insights

provides greater confidence in the findings. The findings from the learner interviews should be interpreted in light of these strengths and limitations.

3. Impact evaluation results

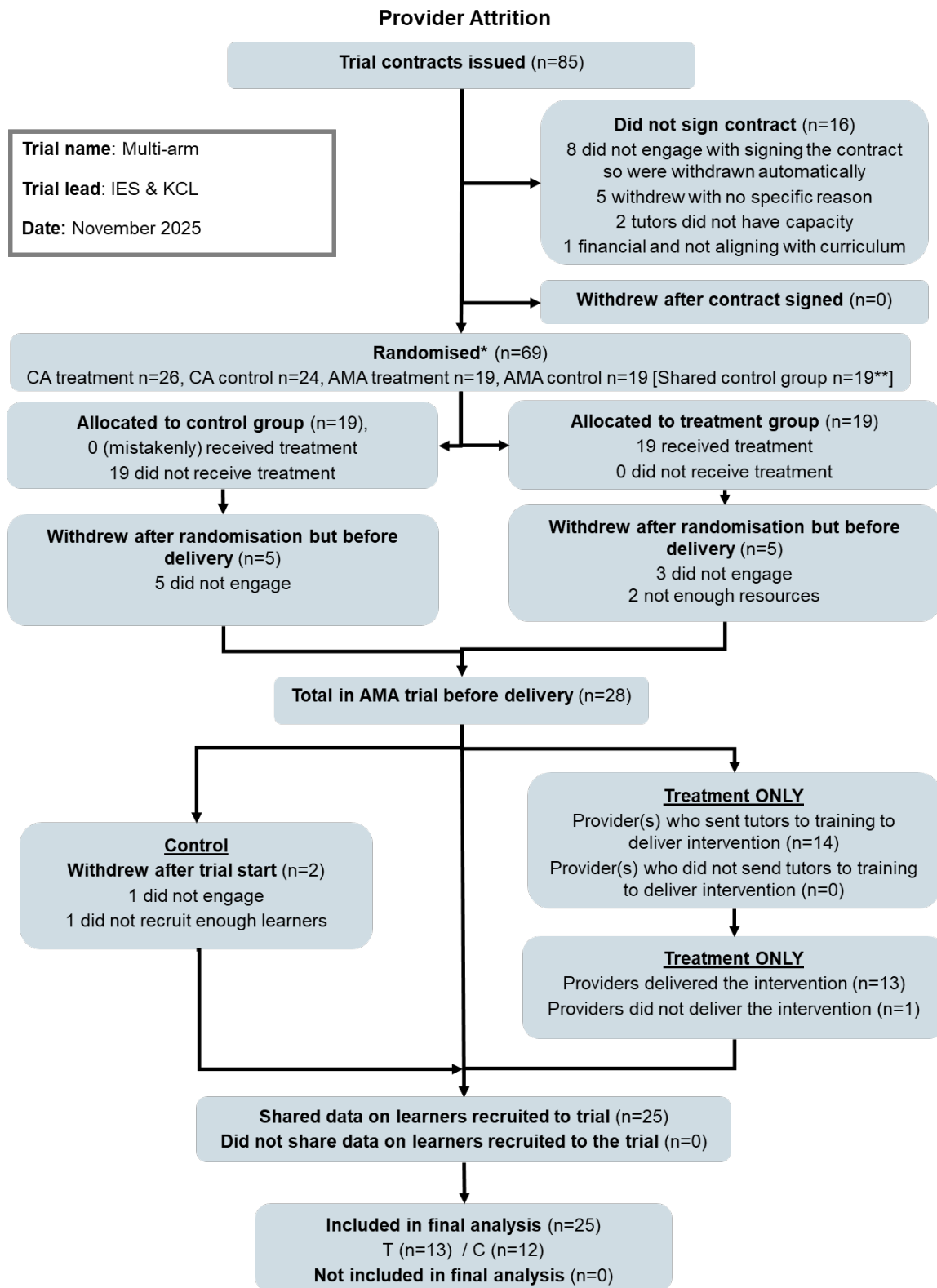
3.1. Participants and attrition

Providers were drawn from a wide range of English regions, including the East Midlands, East of England, London, North West, South East, West Midlands, North East, and Yorkshire and the Humber.

A CONSORT¹¹ diagram of the recruitment stages is shown in Figure 2. A total of 38 providers were randomised to the AMA trial (19 to the control group and 19 to the treatment group). Of these, 13 withdrew after randomisation but before submitting any learner data (7 from the control group and 6 from the treatment group), leaving a total sample size of 25 providers for analysis (12 in the control group and 13 in the treatment group). Overall, reasons for withdrawal included lack of engagement, insufficient resource or not enough learners recruited.

¹¹ CONSORT (Consolidated Standards of Reporting Trials) is an internationally recognised framework for reporting randomised controlled trials. The CONSORT flow diagram shows the progress of all participants through the trial, including enrolment, allocation, intervention, and analysis stages.

Figure 2: Provider attrition map



* Figures below 'Randomised' relate specifically to the AMA trial

** CA control contains 19 from shared control and 5 from CA only control. AMA control includes all 19 from shared control



Table 5 summarises the distribution of providers across regions by treatment allocation. The treatment and control groups both included providers from multiple regions, providing broad geographic coverage consistent with stratification during randomisation. Given the relatively small number of providers in the trial, some regional imbalance was expected and was addressed analytically by including region fixed effects.

Table 5: Provider region

Region	Treatment (at analysis)	Treatment (at randomisation)	Control (at analysis)	Control (at randomisation)
East Midlands	2	3	1	2
East of England	3	4	3	4
London	4	5	2	6
North West	1	2	2	1
South East	1	1	1	1
West Midlands	0	0	0	1
North East	0	1	1	1
Yorkshire and the Humber	2	2	2	2
South West	0	1	0	1
Total	13	19	12	19

Source: ILR. Base: N/A.

Table 6 presents summary statistics for the primary and secondary outcomes at baseline, by treatment status and overall. At endline, 39% of the full sample passed the FSQ Level 1 maths exam, with a higher proportion in the control group (49%) than in the treatment group (29%). For self-reported confidence in maths, mean scores were similar across groups at baseline, indicating good balance prior to the intervention. The composite maths confidence measure showed a similar pattern, with baseline means closely aligned across groups and endline means slightly higher overall.

Table 6: Baseline outcome summary statistics

Outcome	Time	N (A)	Mean (A)	SD (A)	N (C)	Mean (C)	SD (C)	N (T)	Mean (T)	SD (T)
Confidence in maths	Base	221	3.56	1.13	114	3.54	1.08	107	3.57	1.18

Outcome	Time	N (A)	Mean (A)	SD (A)	N (C)	Mean (C)	SD (C)	N (T)	Mean (T)	SD (T)
Confidence in maths (composite)	Base	219	3.87	0.81	112	3.87	0.82	107	3.87	0.80

Note: Columns marked (A) show results for all respondents, columns marked (C) show results for control respondents only, columns marked (T) show results for treatment respondents only.

Source: ILR and baseline learner surveys.

Base: included in 'N' column for each subgroup.

Table 7 shows the characteristics of learners who took part in the trial. Most learners came from areas of medium or low levels of deprivation (68%), with 31% from the most deprived areas – a higher proportion than in the general population but consistent with previous estimates of FE learners' deprivation levels (DfE, 2019). High deprivation was defined as learners in the bottom quintile (20% most deprived areas in England). Around 54% of learners were employed at baseline, while 31% were unemployed. The sample was ethnically diverse, with White learners representing 48% of the sample and substantial representation from Black (25%) and Asian (16%) ethnic groups. Around 7 in 10 learners (71%) were female, and around 20% of participants reported an illness or disability at baseline.

Statistical tests were conducted comparing the sample's demographic characteristics against national FE population benchmarks (DfE, 2025a; DfE, 2025b).¹² This showed that learners in the trial were different from typical FE learners across England in terms of ethnicity and sex. White learners made up a smaller proportion of the trial (50%) than they do nationally (72%), while there were more Black learners (26% vs 8%) and Asian learners (17% vs 11%).¹³ The sample was also more female than the national FE population (72% vs 63%).¹⁴ The proportion reporting ill health was similar to the national rate of 20%,¹⁵ though these measures are not directly comparable and should be treated carefully.

Information on baseline maths confidence was missing for a large proportion of the learners in the trial (76%). Among those who did respond to the survey, most reported being fairly or very confident. This suggests potential selection bias, as those less

¹² Observations with missing values on the relevant variable were excluded.

¹³ $\chi^2(4)=466.69$, $p<.001$

¹⁴ $\chi^2(1)=34.64$, $p<.001$

¹⁵ $\chi^2(1)=1.40$, $p=.236$

confident may have opted out of the trial or the course altogether. Treatment assignment was broadly balanced across both arms.

Table 7: Sample proportions by baseline category

Variable: Term	N	Percentage
Deprivation: Missing	10	[u]
Deprivation: Medium/Low	621	68%
Deprivation: High	283	31%
Employment: Missing	140	15%
Employment: Not Employed	283	31%
Employment: Employed	491	54%
Ethnicity: Missing	32	4%
Ethnicity: White	441	48%
Ethnicity: Mixed/Multiple ethnic groups	28	[u]
Ethnicity: Asian or Asian British	146	16%
Ethnicity: Black, Black British, Caribbean or African	233	25%
Ethnicity: Other ethnic group	34	4%
Sex: Missing	11	[u]
Sex: Male	253	28%
Sex: Female	650	71%
Ill health: Missing	11	[u]
Ill health: Unknown	74	8%
Ill health: No	647	71%
Ill health: Yes	182	20%
Region: East Midlands	115	13%
Region: East of England	338	37%
Region: London	372	41%
Region: North East	[c]	[u]
Region: North West	[c]	[u]

Variable: Term	N	Percentage
Region: South East	28	[u]
Region: Yorkshire and the Humber	33	4%
Maths confidence (baseline): Missing	693	76%
Maths confidence (baseline): Not at all confident; or Not very confident	45	5%
Maths confidence (baseline): Neither confident nor not confident	40	4%
Maths confidence (baseline): Fairly confident; or Very confident	136	15%
Maths confidence (endline): Missing	713	78%
Maths confidence (endline): Not at all confident; or Not very confident	12	[u]
Maths confidence (endline): Neither confident nor not confident	16	[u]
Maths confidence (endline): Fairly confident; or Very confident	173	19%
Treatment arm: Control	444	49%
Treatment arm: Treatment	470	51%

Source: ILR, baseline and endline learner surveys, randomisation data. Base: n = 914.

Some shorthand used in this table. [u] indicates that data are not reported because the unweighted base is fewer than 30. [c] indicates that data has been suppressed to protect anonymity.

Baseline balance was assessed by testing whether treatment and control groups had similar characteristics at the start of the trial. The analysis accounted for region in all cases and adjusted standard errors for the provider level. The results indicate no statistically significant differences between the treatment and control groups on observed baseline characteristics. Across demographic variables (sex, ethnicity and age), socio-economic indicators (deprivation and employment) and health status, the groups were broadly similar, and no differences were statistically significant.

Overall, the results suggest comparable baseline characteristics across groups, though Table 7 highlights substantial missing data in some variables that may have affected the precision of the analyses presented in this section. Table 8 shows the results of the baseline balance checks across key covariates.

Table 8: Balance of baseline characteristics

Covariate	Term	Estimate	SE	P-value	N
Sex	Female	-0.030	0.05	0.537	903
Ethnicity	Mixed/multiple ethnic groups	-0.059	0.07	0.378	882
Ill health and disability	Yes	-0.014	0.03	0.672	829
Deprivation	Deprived	-0.014	0.06	0.827	904
Employment	Employed	0.005	0.08	0.956	774
Age	N/A	0.002	0.001	0.091	914
Confidence in maths	Composite measure	0.002	0.04	0.957	219
Confidence in maths	Main measure	-0.017	0.03	0.577	221

Source: ILR, baseline and endline learner surveys. Base: included in sixth column for each variable. Variation in N is due to exclusion of missing observations.

Taken together, these results suggest that the randomisation achieved good balance on observed baseline characteristics, given the constraints of the trial. The balance analysis also showed no statistically significant differences between the treatment and control groups on baseline confidence in maths. The estimated differences in baseline confidence (both the main and composite measures) were small, imprecisely estimated, and not statistically significant. This supports the conclusion that the two groups were comparable prior to the intervention.

During the trial, 339 of the 914 participants (37%) withdrew from their course, did not complete it, or did not have a registered pass/fail status in the ILR, so their attainment outcome was recorded as failed (0).

For the survey outcomes, Table 9 shows the level of attrition from baseline to endline for each of the confidence in maths measures. For the main confidence measure, only 74 of the 221 learners who completed the baseline survey also responded at endline, corresponding to an attrition rate of 67% for the matched survey sample. Attrition rates were similar across the treatment (69%) and control (64%) groups. The composite confidence measure showed similar attrition patterns.

The similarity of attrition rates across the treatment and control groups suggests that attrition is unlikely to be systematically related to treatment assignment. However, the high overall level of attrition substantially reduced the effective sample sizes for the secondary outcomes. This limited the study’s statistical power and increased uncertainty around the estimates. Further exploration of the mechanisms of missingness and the use of multiple imputation is described in section 3.2.4.

Table 9: Attrition – secondary outcomes

Outcome	N (Baseline)	N (Endline)	Attrition rate	Attrition rate (Control)	Attrition rate (Treatment)
Confidence in maths	221	74	67%	64%	69%
Confidence in maths (composite)	219	73	67%	64%	69%

Source: baseline and endline learner surveys. Base: included in second and third column for each measure.

3.2. Outcomes and analysis

3.2.1. Primary analysis

The primary analysis used a multilevel logistic regression model to estimate the ITT effect of the intervention on FSQ Level 1 maths attainment (pass/fail). Table 10 shows the estimated treatment effect, which was not statistically significant ($d=.003$, $p=.995$), with the 95% confidence interval of the log-odds estimate spanning from -0.87 to 0.88 . This indicates high uncertainty around the estimate.

The findings show an increase in the probability of passing FSQ Level 1 maths of approximately 0.1 ppt for those in the treatment group (based on a log-odds coefficient of 0.003, a control probability of 51.1%, and a predicted probability for the treated of 51.2%). In risk ratio terms, this represents a 0.2% increase in the probability of the treatment group passing relative to the control, effectively indicating a null result.

Table 10: Results – primary outcome (attainment)

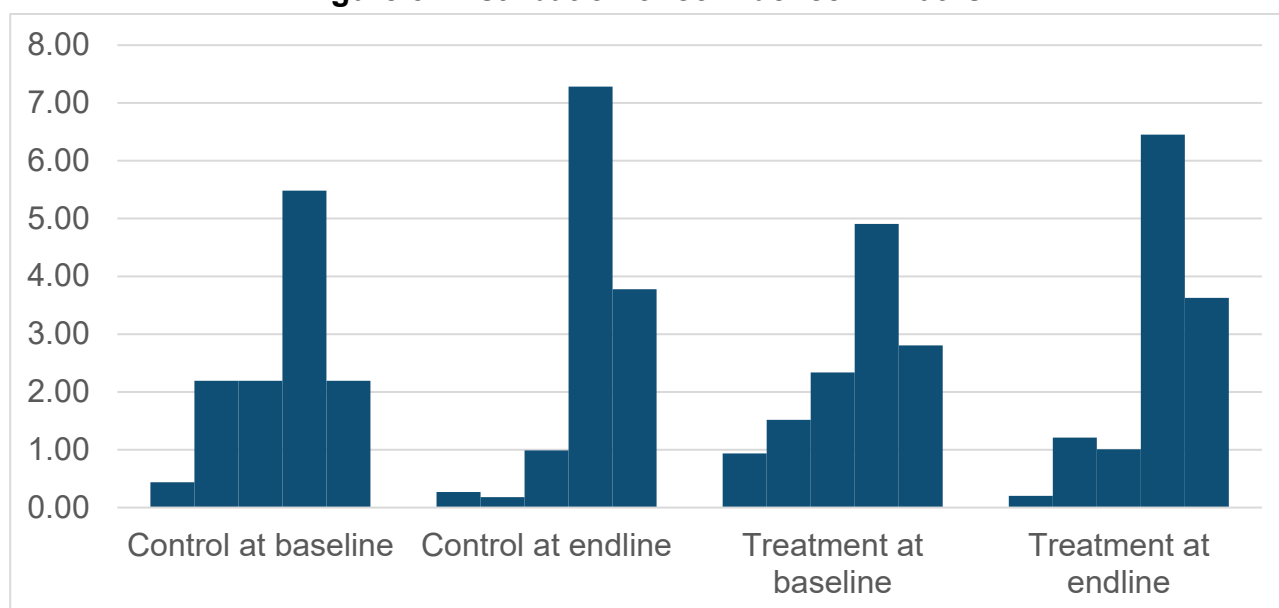
Outcome	Estimate	SE	CI Lower (95%)	CI Upper (95%)	P-value	Predicted probability (control)	Predicted probability (treatment)
Attainment	0.003	0.45	-0.87	0.88	0.995	51.1%	51.2%

Source: ILR. Base: n = 734.

3.2.2. Secondary analysis

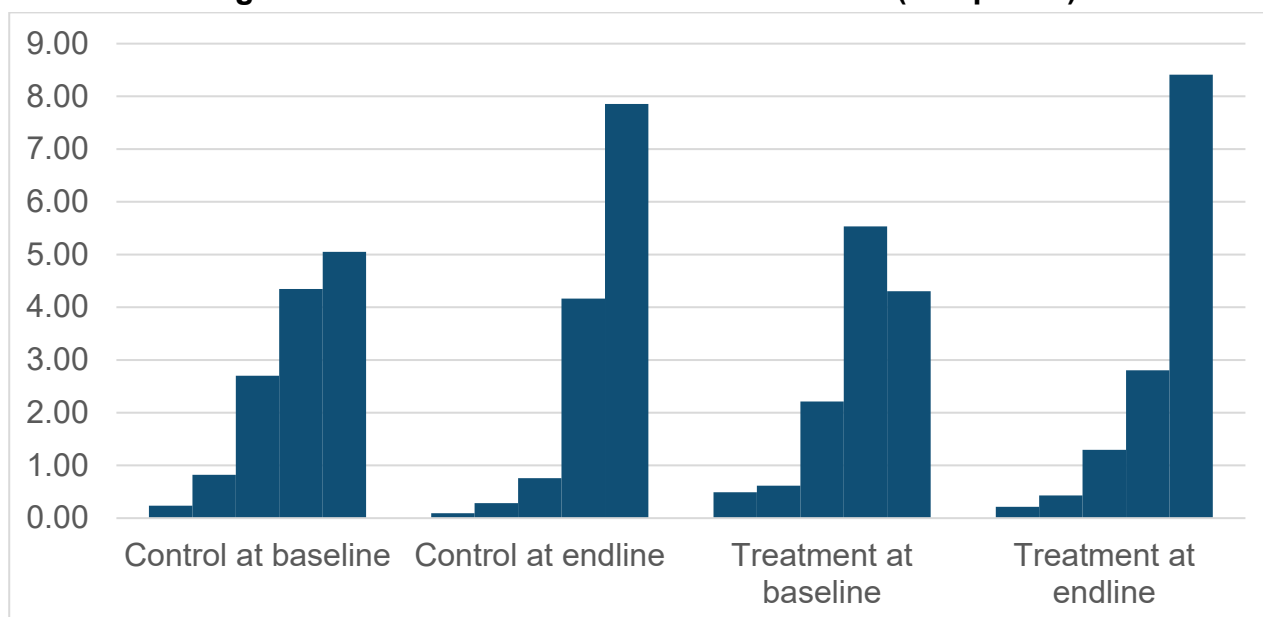
For the single confidence in maths measure, the estimated effect was negative and not statistically significant at the 95% confidence level ($d=-.01$, $p=.956$). In contrast, for the composite measure of confidence in maths, the estimated treatment effect was positive, and approached statistical significance at the 95% level ($d=.27$, $p=.066$). These results are shown in Figure 3 and Figure 4.

Figure 3: Distribution of confidence in maths



Source: baseline and endline learner surveys. Base: n = 62 - 139.

Figure 4: Distribution of confidence in maths (composite)



Source: baseline and endline learner surveys. Base: n = 61 – 139.

These results appear contradictory given that both measure aspects of maths confidence. However, the estimates for each individual question within the composite score (Table 12) help explain this pattern. The single measure asks about numbers in everyday life (*Overall, how confident do you feel working with numbers in everyday life?*). The results of the individual questions in the composite score show that the positive effect was entirely driven by responses to question a, which was also statistically significant at the 99% confidence level, and question e, which was significant at the 90% confidence level.

For both measures, it is important to emphasise that the small sample size at endline significantly reduced the effective sample size used in the analysis model. This reduction in sample size suggests a substantial loss of statistical power, meaning that the results are highly sensitive to variation.

Table 11: Results – secondary outcomes (confidence in maths)

Outcome	Estimate	SE	CI Lower (95%)	CI Upper (95%)	P-value	N
Confidence in maths	-0.01	0.18	-0.36	0.34	0.956	72
Confidence in maths (composite)	0.27	0.15	-0.02	0.56	0.066	71

Source: baseline and endline learner survey. Base: included in seventh column for each measure.

Table 12: Results for individual questions in composite confidence in maths measure

Question (How confident do you feel about doing the following things in everyday life?)	Estimate	SE	CI Lower (95%)	CI Upper (95%)	P-value	N
a) Checking your change is right when you have bought something.	0.53	0.13	0.28	0.78	0.000	67
b) Working out the best deals when shopping	0.12	0.19	-0.25	0.49	0.512	69
c) Helping children with maths homework or talking about maths/numbers with children	-0.38	0.51	-1.38	0.62	0.456	39
d) Understanding interest rates on bank statements	0.38	0.26	-0.14	0.90	0.149	69
e) Keeping track of your bank account balance	0.20	0.11	-0.02	0.42	0.079	71
f) Working with numbers as part of a job	-0.20	0.25	-0.68	0.28	0.417	37

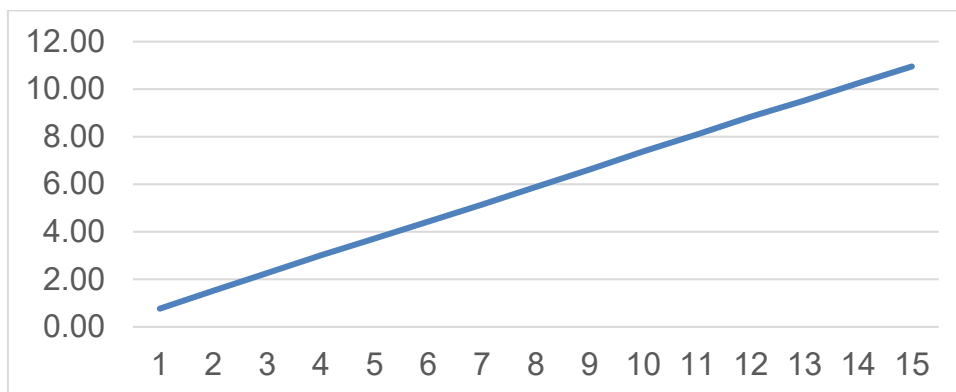
Source: baseline and endline learner survey. Base: included in seventh column for each sub-question.

Additional [data tables](#) contain more detail on survey responses.

3.2.3. Compliance analysis

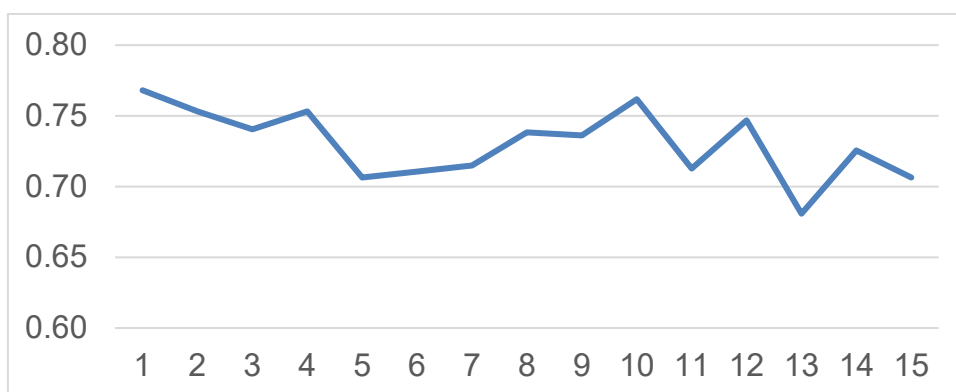
Compliance was measured by the total number of sessions each learner attended fully or partially. Only 233 out of 470 treated learners met the required compliance threshold of 80% (attended 12 or more out of 15 sessions), so compliance-adjusted analysis was conducted. Figure 5 shows the average cumulative number of sessions attended by the treatment group (averaged at each week across the 15-week intervention) and Figure 6 shows the proportion of participants attending each session by week.

Figure 5: Attendance (average cumulative)



Source: MSS monitoring data. Base: n = 470 (non-responses are excluded).

Figure 6: Attendance (by session)



Source: MSS monitoring data. Base: n = 470 (non-responses are excluded).

The CACE estimate showed a positive effect on the primary outcome, with learners who complied showing a 3.3 ppt higher pass rate than the ITT estimate. Although the p-value for the CACE estimate was smaller than in the primary analysis, the estimate remained not statistically significant ($p=.782$).

For confidence in maths, the results remained largely the same, both in terms of point estimates and significance levels. These results suggest that adjusting for compliance did not meaningfully alter the estimated effects of the intervention on either the primary or secondary outcomes.

Table 13: Results – CACE

Outcome	Estimate	SE	CI Lower (95%)	CI Upper (95%)	P-value	N
Attainment	0.03	0.12	-0.20	0.27	0.782	734

Outcome	Estimate	SE	CI Lower (95%)	CI Upper (95%)	P-value	N
Confidence in maths	-0.01	0.21	-0.42	0.39	0.954	72
Confidence in maths (composite)	0.32	0.19	-0.04	0.69	0.079	71

Source: ILR, baseline and endline learner survey. Base: included in seventh column for each outcome measure.

3.2.4. Missing data analysis

Baseline characteristics did not correlate with missing data, except for certain providers. However, the correlations varied in direction, some providers showed positive correlations while others showed negative correlations (see Appendix 6). Importantly, there was no difference in attrition between the treatment and control groups across all outcomes.

For baseline data alone, as shown in Table 14, missing data varied greatly between variables, with higher levels in the confidence in maths survey outcome, where over three-quarters of observations were incomplete.

Table 14: Missing data – baseline covariates

Variable	Complete	Incomplete	Imputed
Confidence in maths	221	693	612
Confidence in maths (composite)	219	695	614
Ethnicity	882	32	19
Deprivation	904	[u]	[u]
Employment	774	140	69

Source: ILR, baseline and endline learner surveys. Base: n = 914.

Some shorthand used in this table. [u] indicates that data are not reported because the unweighted base is fewer than 30.

Missing data were handled using multiple imputation by chained equations (MICE), following the recommendations of White et al. (2011) and Carpenter & Smink (2020). A total of 10 imputation datasets were generated with imputation methods tailored to each variable's distribution. Linear regression was used for continuous variables and logistic regression for binary variables.

The imputation model included all main analysis variables including treatment, age, sex, ill health, and the 2 outcomes of maths confidence at baseline. Provider and region were also included as covariates in the imputation model and augmented regression was used where necessary to avoid convergence issues. Results were pooled across imputed datasets using Rubin's rules via Stata's "mi estimate" command.

Table 15 shows that the imputation-based analysis of the primary outcome (attainment) resulted in an estimated treatment effect that was marginally larger than the complete-case estimate, showing a 1 ppt difference due to the intervention. However, this was not statistically significant ($d=.01$, $p=.926$). This suggests that results are not driven by missing data bias, though the imputed estimates remain imprecise.

Table 15: Results – MICE

Outcome	Estimate	SE	CI Lower (95%)	CI Upper (95%)	P-value
Attainment	0.008	0.09	-0.17	0.18	0.926

Source: ILR. Base: n = 823.

3.2.5. Sub-group analysis

This analysis examined whether the treatment effect differed across ethnicity, sex and ill health/disability status by testing for interactions between treatment assignment and each subgroup characteristic for the primary outcome, as shown in Table 16. All analyses were underpowered and exploratory, as specified in the protocol.

Across all subgroup analyses, there was no statistically significant evidence that the treatment worked differently for different groups. For ethnicity, although the point estimates suggest a larger treatment effect for non-White groups, the confidence intervals were wide and included zero, indicating substantial uncertainty. Therefore, there was no reliable evidence that the impact of the intervention differed by ethnicity.

For sex, the interaction term for females compared to males was positive (0.65) but not statistically significant ($p=.327$). Similarly, the interaction term for those with ill health or disability was around 0.5 and equally imprecise.

Table 16: Sub-group effects of treatment interactions

Variable	Term	Estimate	SE	CI Lower (95%)	CI Upper (95%)	P-value
Ethnicity	White ethnic group	0.91	0.98	-0.17	0.41	0.354
Sex	Female	0.65	0.66	-0.65	1.94	0.327
Ill health/ disability	Yes	0.48	0.55	-0.59	1.55	0.379

Source: ILR. Base: n = 734.

3.2.6. Additional analyses and robustness checks

As a robustness check, the primary model was re-run excluding observations that did not have an outcome instead of imputing fails. As expected, the effect became larger and approached statistical significance ($d=.74$, $p=.084$) compared to the main analysis ($d=.003$, $p=.995$). The log-odds of 0.74 implies a 15 ppt increase in the probability of a pass compared to the control group (probabilities for treatment: $p_1=71\%$, and control: $p_0=56\%$) which corresponds to a Cohen's $h=0.31$. This suggests potential selection bias when excluding learners who did not have a pass/fail outcome for their FSQ Level 1 course from the analysis (rather than including them as having failed), likely because learners who are most likely to pass are also more likely to take the final exam.

Table 17: Results – robustness check

Outcome	Estimate	SE	CI Lower (95%)	CI Upper (95%)	P-value
Attainment (model 1)	0.74	0.43	-0.10	1.58	0.084

Source: ILR.
Base: n = 549.

3.2.7. Estimation of effect size

Effect size results for the primary outcome of attainment (Cohen's h) showed a near-zero effect of 0.001 standard deviations. The effect sizes for confidence in maths were calculated using the square root of the total variance. The single item confidence in maths measure showed a very small negative effect size of -0.02 standard deviations, while the composite measure showed a much larger effect size of 0.52 standard deviations.

Table 18: Results – conditional effect sizes

Outcome	Effect Size	ES CI Lower (95%)	ES CI Upper (95%)	N
Attainment (Cohen's h)	0.001	-0.10	0.10	734
Confidence in maths (Hedge's g)	-0.02	-0.67	0.64	72
Confidence in maths (composite) (Hedge's g)	0.52	-0.03	1.07	71

Source: ILR, baseline and endline learner surveys.
Base: included in fifth column for each outcome measure.

For the unconditional estimates, an “empty” model (a model without any additional covariates) was used, which yielded predicted baseline probabilities of 45% for the control group (p_0) and 48% for the treatment group (p_1). These estimates show a difference of 2.5 ppt with a Cohen's h of 0.05, or a 6% relative increase in the probability of passing for the treatment group. An empty model was also used to estimate unconditional cluster and residual variance for the confidence in maths effect sizes. Table 19 shows the effect size (Cohen's h or Hedge's g) and the corresponding 95% confidence intervals for each outcome.

Table 19: Results – unconditional effect sizes

Outcome	Effect Size	ES CI Lower (95%)	ES CI Upper (95%)	N
Attainment (Cohen's h)	0.05	-0.04	0.14	734
Confidence in maths (Hedge's g)	-0.01	-0.42	0.40	72
Confidence in maths (composite) (Hedge's g)	0.38	-0.03	0.78	71

Source: ILR, baseline and endline learner surveys.
Base: included in fifth column for each outcome measure.

3.2.8. Estimation of ICC

The conditional ICC for attainment was 0.2, suggesting that learners within the same providers had somewhat similar outcomes (Table 20). In contrast, the model estimates for the maths confidence outcomes indicate effectively no clustering at provider level after accounting for covariates, with the estimated ICCs at effectively 0. The

unconditional ICCs showed higher levels of similarity within providers when not controlling for baseline covariates (Table 21).

Table 20: ICC conditional estimates

Outcome	Cluster Variance	Residual Variance	ICC	N
Attainment	0.78	3.29	0.19	734
Confidence in maths	0.00	0.29	0.00	72
Confidence in maths (composite)	0.00	0.27	0.00	71

Source: ILR, baseline and endline learner surveys.
Base: included in fifth column for each outcome measure.

Table 21: ICC unconditional estimates

Outcome	Cluster Variance	Residual Variance	ICC	N
Attainment	0.99	3.29	0.23	908
Confidence in maths	0.09	0.66	0.11	201
Confidence in maths (composite)	0.08	0.43	0.16	200

Source: ILR, baseline and endline learner surveys.
Base: included in fifth column for each outcome measure.

Given the post-attrition sample of 25 providers, with an average cluster size of 37 learners, baseline-endline correlation of 0.2 and an intracluster correlation (ICC) of 0.19, this pilot trial was powered to detect a difference of 14 ppt from a baseline, or an implied Cohen’s h (MDES)¹⁶ of 0.51.

¹⁶ The Minimum Detectable Effect Size (MDES) is the smallest difference (effect) between a treatment group and control group that a study has enough statistical power to detect as statistically significant.

IPE results

This section examines the delivery of the Adapted Mastery Approach programme as experienced by learners, tutors, and providers, using both qualitative and quantitative data from interviews, surveys and attendance records. The IPE covered tutor and learner engagement, delivery practices and contextual factors that affected participants. The aim of the IPE analysis was to investigate central features of trial delivery, including compliance, attrition, fidelity, feasibility and readiness for future trials, with recommendations for future implementation. Key IPE findings are provided below, with further detail provided throughout the chapter.

- **Compliance:** There was a high level of compliance among tutors with the training requirements and delivery of intervention sessions. Among tutors where training data was available, most (83%) attended all 4 training sessions while the remaining attended 2 or 3 sessions out of 4, and almost all tutors (95%) reported delivering each of the 15 sessions to learners. Among learners, 50% met the requirement of attending at least 80% of intervention sessions to be considered compliant with the intervention.
- **Attrition:** A relatively high number of providers dropped out of the trial post-randomisation and pre-training.
- **Fidelity:** Tutors largely following the intervention lesson plans, though many needed to make small adaptations to account for learners' different needs, mode of delivery, and gaps in coverage of the curriculum. The evidence suggests that the intervention does differ from BAU teaching approaches, and contamination is unlikely to have been an issue.
- **Feasibility:** Overall, the findings suggest the intervention is practical and feasible to deliver, though some barriers and potential improvements were identified, including adaptations to the teaching resources, reducing the training burden, and making the course more appropriate for online delivery.
- **Readiness for trial:** The pilot evaluation findings imply that the intervention has a reasonable level of readiness and would be suitable for further testing with some adjustments, including improving curriculum alignment, applicability for learners at different levels and online learners, and addressing recruitment and attrition challenges.
- **Perceived outcomes:** Treatment tutors reported positive outcomes from their experiences of delivering the intervention, including improved teaching practices, shifts in their pedagogy and mindset, and professional development. Learners in the treatment group reported improved confidence, willingness to attempt maths problems, and application of maths in everyday contexts. However, it should be noted that positive impacts were also reported prominently among control group tutors and learners, including in surveys with control group learners.

3.3. Compliance

This section describes findings related to compliance, including whether tutors took part in training and support, whether they delivered each of the sessions, and whether learners attended sessions.

There was relatively high compliance with the training requirements among tutors who participated in interviews. Monitoring data collected by the MSS showed that, for the 29 tutors where data was available, 24 (83%) of these attended all 4 training sessions. A further 3 tutors completed 3 out of 4 sessions, while the remaining 2 completed 2 out of 4 sessions. All tutors interviewed had attended at least 3 of the 4 training sessions, with most tutors having attended all 10 hours of training.

Twenty staff (69%) attended at least 8 out of 12 lesson study sessions, while 4 attended between 1 and 7 sessions, and 5 did not attend any sessions. However, 3 of those who did not attend any were characterised as delivery staff rather than tutors in the monitoring data. They had attended training to support tutors in their college to deliver the programme, rather than delivering the programme themselves.

Monitoring data was available from 21 tutors about delivery of each of the 15 sessions. All tutors reported delivering all the required sessions, except for 1 who delivered 13 out of 15 sessions. While there may be a bias in responses towards those who delivered the required sessions, this data suggests that there was high compliance with the intervention in terms of delivering the 15 sessions.

A total of 233 learners (50%) in the treatment group attended 12 or more sessions, which was defined as being compliant. The average attendance rate by treatment learners (both compliant and non-compliant) was 11 sessions out of 15, while the average attendance for compliant treatment learners was 14 sessions out of 15. For each session, around three-quarters of treatment group learners attended, and this attendance rate remained fairly consistent throughout the course (see Table 22). No equivalent data was collected for the control group.

Table 22: Attendance rates across sessions in the treatment group

Session name	Number of treatment learners who attended fully/partly	Percentage of treatment learners who attended fully/partly
Session 1: Maths Mindsets and Resilience	361	77%
Session 2: Graphs and Charts	354	75%
Session 3: Averages and Range	348	74%
Session 4: Direct Proportion	354	75%
Session 5: Scales, Maps and Units	332	71%
Session 6: Multiplication, Division and Estimation 1	334	71%
Session 7: Multiplication, Division and Estimation 2	336	71%
Session 8: Fractions of Amounts	347	74%
Session 9: Ratio	346	74%
Session 10: Fractions, Decimals and Percentages	358	76%
Session 11: Percentages	335	71%
Session 12: Probability	351	75%
Session 13: Perimeter	320	68%
Session 14: Area	341	73%
Session 15: Volume and capacity	332	71%

Source: MSS monitoring data.

Base: n = 470 (non-responses are excluded).

Note: Results show what percentage of everyone who answered the question gave 1 of the positive response options (“Attended all”, “Attended part”).

Two providers in the treatment arm also took part in another Adult Numeracy Trial, Embedded Maths. Learners from these providers were excluded from the impact and

survey analysis, and were not approached for interviews for the AMA IPE to avoid duplication and reduce participant burden.

3.4. Attrition

Monitoring data collected by the MSS provided information about the level of and reasons for attrition from the trial, in both the treatment and control group. This is also summarised in the attrition map (see Figure 2).

3.4.1. Treatment

In total, 19 providers were randomised to the AMA treatment group, though 6 withdrew soon after randomisation (32% attrition rate in treatment group). Of these, 2 did not engage in any trial activities, including training, and withdrew due to non-compliance. Another 2 felt they no longer had the required resources to engage in the intervention; 1 because changes to trial set-up, training and study session requirements had increased tutor commitment, and 1 because they had lost a tutor and could no longer commit to delivering the course to the required number of learners. The final provider felt, after receiving the training, that AMA delivery did not align with their current delivery model, especially because they delivered shorter and more intensive courses that were adaptable to a variety of levels. The provider reported that some of their tutors might be incorporating skills they learned during the initial training, but not necessarily as prescribed.

3.4.2. Control

In total, 19 providers were randomised to the control group, though 7 withdrew early after randomisation (37% attrition rate in the control group). Of these, 5 did not engage in any trial activities and withdrew due to non-compliance. The remaining 2 could not recruit enough learners to deliver the required course.

Providers in the control group who were interviewed as part of the IPE expressed disappointment at being allocated to the control group. However, some found it somewhat of a relief not to have to implement the intervention on the short timelines between sign-up and the intervention start date. Most of these providers also stated that they understood the importance of a control group and still felt they were making a valuable contribution to the research.

Interviews conducted by Ipsos with withdrawn providers across all the Adult Numeracy Trials suggested that the specific reasons for withdrawing from the AMA trial, as opposed to general reasons for dropping out of trials, were that the training requirements were seen as too burdensome, especially for providers with tutors who only worked part time.

3.5. Fidelity

This section describes the extent to which the intervention was delivered as intended by treatment providers. It explores key assumptions identified in the Theory of Change (Table 1) by providing evidence on the training and support delivered to tutors and their uptake, the subsequent teaching delivered to learners, and the extent to which this teaching reflected the intended approach.

3.5.1. Tutor training and support

There was high attendance at the pre-delivery training among tutors. Among the 27 tutors where data was available, 22 attended all sessions, and the remainder attended either 2 or 3 sessions. Among tutors interviewed, some reported missing the final session for various reasons including pre-booked annual leave and illness. Several tutors who had missed one of the sessions highlighted that they had reviewed the materials for that training session or spoken about the content of the session with a colleague who had attended.

Most tutors interviewed attended the lesson study sessions, but there was higher variation in the rate of attendance than for the pre-intervention training. Tutors interviewed also reported varied levels of engagement in their lesson study sessions, with those who had a highly engaged group tending to find the sessions more beneficial than those whose groups were less focused on discussion.

Overall, these findings suggest a relatively high level of engagement with training and support among tutors.

3.5.2. Implementation of approaches and activities

Many tutors who participated in interviews largely followed the intervention lesson plans, following the 15-lesson structure and using the resources provided. However, some tutors spoke of having to make adjustments to account for different lesson lengths or frequencies. Many tutors interviewed reported needing to make minor adaptations to account for learners' different needs, mode of delivery, and missed areas of the curriculum, while a small number of tutors made more significant adaptations. The intention was for tutors to deliver the lessons without adaptations, so these changes represent a deviation from the intervention as planned.

The extent of the adaptations varied between contexts, with some tutors implementing only minor changes while others felt the need to make more significant changes to the lesson plans and resources. Many of the tutors interviewed who made adaptations for learners who spoke English as a second language or had SEND found minimal adaptations were often adequate. Some provided examples of these small changes, such as altering the background colour of the slides for a learner with dyslexia or

replacing ratio tables with double number lines for a learner with autism who struggled with the non-linear nature of ratio tables.

Adaptations varied more when made for learners with lower maths skills. To account for different skill levels, some tutors described making relatively minimal adaptations to the volume of content for the lesson plans. Some felt that the lesson slides had too much content, and covering all the material in a single lesson was very rushed and did not give learners enough time to absorb the information. Conversely, others felt that some lesson presentations lacked sufficient examples or practice questions, so they added additional ones. These contrasting adaptations – both made to account for learners with lower maths skills – highlight that a one-size-fits-all approach may not be appropriate for settings as varied as adult education providers.

A small number of tutors interviewed reported only implementing some elements of the intervention and choosing not to deliver elements they felt were not useful for their setting. This represents a substantial deviation from the intervention and indicates low fidelity to the programme design in these settings. One tutor interviewed described using the intervention resources as a “toolbox” and selecting interesting aspects rather than following the programme precisely. Another felt there was no pressure to deliver the course as written.

“When we said [to the trainers] we couldn’t deliver it in this order, or on this date, then it was very much, ‘Do it when you can, and do it how you can, and we’re happy if you do anything’, type of thing.” – Tutor, treatment group

Several tutors interviewed reported receiving advice by trainers that encouraged them not to comply with the intervention’s lesson resources, such as making adaptations to slides or adding content to cover topics not included in the intervention. This was confusing for tutors and shows that trainers were not sufficiently enforcing the intervention requirements. This likely explains at least some of the adaptations made by tutors as they were not receiving consistent messaging about the need to deliver the intervention as written, without deviation.

“I queried whether this would affect the results of the trial, or not. Because it wasn’t exactly following the mastery programme in the same way as if it would have been if I hadn’t had to put that extra session in. But I was told to continue doing it, so I did.” - Tutor, treatment group

Conversely, some providers and tutors interviewed spoke about senior staff enforcing programme fidelity, suggesting that buy-in from managers helped increase fidelity to the intervention’s design.

“Some of the ordering in terms of the curriculum order and the way that the topics were taught were sometimes questioned [by tutors] ... my message to my staff was, ‘Well, we’re testing this approach ... We have go with that to see if it does

have a positive impact for learners'. So, I was very keen for them to keep it in the order, and obviously, in the design that had been sent over and designed in that way." - *Provider, treatment group*

In the endline survey, treatment learners surveyed reported a high level of use of mastery approaches in their course, including helping them see how different maths topics connect and encouraging use of clear mathematical language to explain their thinking. However, control learners also reported high use of these strategies. Both treatment and control learners also reported high use of the approaches used in the Contextualised Approach trial arm, such as using real life examples to think about maths. That the learner survey findings suggest high use of all approaches across all arms, despite interviews with tutors not finding any evidence of contamination, suggests that it may be difficult for learners to assess differences in teaching strategies.

3.5.3. BAU and contamination

Control group's experience of BAU and contamination

Control tutors delivered FSQ Level 1 Maths courses without the AMA intervention. Although control tutors interviewed described teaching practices that resembled aspects of maths mastery, this does not indicate intentional use of the AMA intervention. Tutors typically described only using 1 or 2 mastery principles or approaches and not in the systematic way required by AMA. Only one control tutor interviewed reported being aware of and using mastery approach. Several reported having heard of mastery or encountering it in previous roles but not applying it in their current teaching, and the rest had no formal awareness of mastery approaches. Further, there was no indication that any of the control tutors interviewed had gained access to the AMA lesson plans and resources used by treatment tutors

However, many control tutors described some elements of maths mastery principles or approaches when discussing their teaching practices, even though they reported no awareness of mastery. In particular, control tutors spoke about building deep understanding of concepts before moving on to the next topic, focusing on core skills and their application in different contexts, encouraging learners to discuss concepts and use mathematical language, and building learner confidence in their ability to learn maths. This suggests that some elements of the mastery approaches may have been present in the control lessons, but not in the structured way required by AMA or using the AMA resources.

Learner interviews revealed that the teaching approaches described were not notably different between treatment and control learners. Both treatment and control learners reported that their tutors used mastery principles throughout their sessions. The main difference between the teaching styles described by treatment and control learners in interviews was the number of methods taught. AMA learners who participated in interviews reported that they had learned multiple methods to solve the same maths

problem, whereas control learners primarily reported learning one method. Many of the control learners interviewed expressed a desire to learn multiple methods, explaining that if they did not understand the method their tutor taught them, they would search online for alternative solutions.

Treatment group's experience of BAU

Treatment tutors interviewed felt that while there were some similarities between the AMA approach and their BAU practices, certain elements of the intervention were new. The key differences highlighted were: the ordering of topics; teaching multiple approaches; the use of visual learning practices and discussion; and new approaches such as ratio tables.

3.6. Feasibility

This section explores to what extent the intervention is likely to be practical and feasible. In examining this, the section describes the experiences of providers and tutors in delivering the intervention and their views on perceived outcomes for tutors and learners. It also explores learners' experiences and their perceived outcomes.

3.6.1. Tutors' experiences of delivering the intervention

Feasibility of delivery within the FSQ Level 1 maths course

Tutors interviewed had a largely positive perception of the intervention. They reported that the intervention was easy to deliver within the FSQ Level 1 maths course hours, and many said they enjoyed delivering it. Many were also positive about various elements of the intervention, including the use of concrete materials such as blocks, visual models such as ratio tables, the opportunity to foster more discussion and use mathematical language, and encouraging learners to adopt a growth mindset. Additionally, while some tutors interviewed found that adapting the resources was time-consuming, many appreciated having the lesson materials already developed and ready to deliver.

“I loved it. Absolutely loved it. And so did all the learners. It was good to think about things in a different way. Some of the techniques I use anyway, but having the slides already written that were done so that they were very animated, the learners enjoyed all the animation ... at first, they [the learners] were a little bit wary, but as they got into it, they were really excited by it. Their feedback has been amazing. Their mock results are fantastic ... I'm looking forward to starting it again.” – *Tutor, treatment group*

While the overall view of the intervention was positive, some tutors interviewed highlighted that not everything in the curriculum was included in the 15 lesson plans, requiring them to cover examinable topics in addition to the intervention. This included key skills for exams, such as interpreting pie charts or understanding and applying

Brackets, Indices, Division, Multiplication, Addition, and Subtraction (BIDMAS). Several tutors therefore noted that they had to add content to some of the intervention sessions or cover these concepts in revision sessions to ensure their learners were adequately prepared for the exams. One tutor commented that the reason for these omissions was not explained, which they found frustrating.

Some tutors interviewed also commented that the intervention did not always account for the level of maths their learners were starting with. For example, some tutors questioned the ordering of the intervention lessons, which differed from how they would normally deliver the course. Several tutors commented that they would normally begin with basic number skills such as multiplication and then build to more advanced topics, but the AMA course began with graphs in week 1, a more advanced concept, while multiplication was not covered until weeks 5 and 6.

“I found that a bit challenging, just because I feel in some ways, some of the learners would need to have those [number skills] lessons sooner before doing some of the other topics.” - *Tutor, treatment group*

Feasibility and usefulness of the resources, training and support

There was overall satisfaction with the content and quality of training among tutors interviewed. These tutors reported that the initial training sessions were high quality and useful in preparing them to deliver the intervention. In particular, they enjoyed the interactive nature of the training and appreciated the opportunity to speak with other tutors during the sessions and hear different perspectives. Some tutors and providers who participated in interviews also commented that the training had given tutors an opportunity to explore different approaches to adult teaching which was beneficial beyond the intervention. However, others also commented that the training sessions were quite long and more time-consuming than anticipated, with suggestions that the training could be shortened by moving at a faster pace.

One potential improvement suggested for the training was further exploration of the concepts and intentions behind AMA approaches. While many tutors interviewed found the training helpful and felt they had gained greater clarity on how to teach AMA, some felt that the training had not adequately explored the underlying concepts of AMA and was too focused on practical delivery. Several tutors and providers interviewed suggested having a more in-depth introduction to mastery principles and concepts in the training before moving on to applications.

“I just don’t feel like the tutors ... left really knowing what mastery was, what was the idea behind mastery, what was mastery trying to achieve and instill in learners. I think a lot of people left there still thinking that mastery was about the modelling that was used rather than trying to develop a deeper understanding in learners of maths concepts.” - *Provider, treatment group*

Tutors interviewed had mixed views on the usefulness of the lesson study sessions. Some found it beneficial to be able to talk through different implementation strategies with other tutors and reflect on their own practices, while others felt the sessions were quite repetitive and offered little additional insight since the lesson resources and training slides were already available. The weekly lesson study sessions were also challenging for many tutors with already busy workloads. Several tutors suggested replacing the mandatory weekly sessions with optional drop-in sessions that tutors could attend when they needed support.

“It’s an extra two hours that you’ve got to find that you didn’t have in the first place ... it was a bit challenging.” - *Tutor, treatment group*

The timing and size of the lesson study sessions were raised as factors affecting perceptions of usefulness. Due to the variability in tutors’ timetables, some tutors had already delivered that week’s session while others were still preparing to deliver it, which meant not everyone was able to contribute to the discussions. Regarding size, some tutors interviewed enjoyed being in small groups because it gave them more opportunity to ask questions of the trainer, while others felt that larger groups fostered more discussion across a wider range of experiences. As with the pre-delivery training, some tutors commented that it would have been helpful to have more of a focus on the intentions of AMA approaches during the lesson study sessions.

“What I would have welcomed is the opportunity to say, ‘This is a maths mastery approach, because this is what we want you to get out of this session. This is why we are using this approach.’ I didn’t feel there was enough of that in those sessions.” - *Tutor, treatment group*

Enablers to delivery

While some tutors interviewed found that adapting the resources was time-consuming, many tutors interviewed appreciated having the lesson materials already developed and ready to deliver. The availability of the lesson slides and materials meant that tutors did not have to spend time planning how to implement the strategies they had learned. Though many tutors made slight adaptations to account for the needs of their learners, overall, the pre-prepared resources were seen as making it easy for tutors to integrate the intervention into their courses.

The intervention was broadly seen as encouraging both tutor and learner engagement and motivation. Providers interviewed reported that learning new teaching approaches through the intervention had been motivating for their tutors, and that this contributed to their commitment to the approach.

“She’s [the tutor] really enjoyed delivering it. She found it’s given real purpose and direction to her lessons.” - *Provider, treatment group*

Many tutors interviewed felt that learners were highly responsive to the intervention. They cited several reasons for this, including the visual elements of the intervention, the interactive nature of the sessions, and the confidence-building aspects.

“They love all the different bits and pieces, they like the fact that it’s interactive, they like the fact that it’s very practical, they love the key ideas.” - *Tutor, treatment group*

Several tutors teaching the course over 15 weeks highlighted the benefit of a 15-week course. For example, 1 tutor’s usual course was 8 sessions and quite intensive, and they felt that having learners consistently for 15 weeks allowed them to cover topics in greater depth and led to better outcomes. Another felt that applying the mastery principles consistently over 15 weeks had encouraged learners to be more open and positive compared to teaching a 15-week course as BAU. One tutor who taught a 30-week course delivered the first 15 weeks following AMA and had 2 learners sit and pass early exams after completing the AMA lessons, which they felt demonstrated that the approach was effective.

Barriers to delivery

Though the intervention was designed to be delivered face-to-face, providers were permitted to use AMA online if that was their usual course delivery style. Tutors interviewed who delivered online courses struggled more to integrate the intervention and engage learners. They felt that the intervention resources assumed in-person delivery, requiring online tutors to adapt the resources. However, given the nature of some of the resources, these adaptations could be quite time-consuming. Many also noted that some of the teaching strategies were highly visual or designed to be interactive, which were difficult to deliver online and felt more suited to in-person delivery.

“I think it would've been helpful to know how we were delivering it, because it was written for a face-to-face delivery, and we weren't ever going to be that. We don't deliver like that, and all the activities and the materials were around face-to-face delivery, and we weren't geared for that. So, we had to keep going and finding alternatives or adjusting, and I think the material needed to either cover both options, or it needed to be quite clear how you needed to deliver it.” - *Tutor, treatment group*

Online delivery also influenced learner engagement. Factors such as IT confidence, learners’ willingness to participate in online discussions, and the reduced ability to monitor understanding remotely affected how well some learners could engage, particularly those needing closer support.

Classes with learners who had a wide range of needs also required more adaptations. For example, one tutor spoke of teaching a mixed Level 1 and Entry 3 class, while another had a class with both native English speakers and learners who spoke

English as a second language. Where tutors had to consider a wide range of needs, it was sometimes difficult for them to implement the intervention as planned as elements of the intervention that worked for some of their learners did not necessarily work for others.

Though the lesson materials were generally seen to be of high quality and facilitated tutors' delivery of the intervention, some tutors interviewed highlighted areas of the materials that needed further refinement before wider rollout. These tutors flagged concerns such as errors in the lesson slides, examinable curriculum missing from the intervention content, lesson ordering that did not match learners' pre-existing maths knowledge, and lesson slide packs that were poorly paced. While tutors often made small adaptations to remedy these concerns, for some it influenced their perception of the intervention as a whole.

“I think there’s a little bit of distrust when you get information that you know is not right ... mainly missing content and not covering things.” - *Tutor, treatment group*

A small number of tutors commented on differences in engagement related to learners' age or prior habits. Some older learners were described as falling back on familiar methods, particularly those used previously, while 1 tutor observed that younger learners were more willing to try new approaches.

3.6.2. Learners' experiences and engagement of FSQ Level 1 maths

In the endline survey, learners across the treatment and control groups reported similarly high levels of satisfaction with their FSQ Level 1 maths course. This included satisfaction with the quality of teaching (91% for both treatment and control groups), learning materials (89% vs 91%), length of sessions (93% for both groups) length of course (94% vs 92%), the time of day of lessons (87% vs 89%), and information provided about other courses (91% vs 94%). Overall course satisfaction was rated positively by 96% of treatment learners and 92% of control learners (Figure 7). The difference between the 2 groups was not statistically significant.

Figure 7. Overall, how satisfied or dissatisfied were you with the course?



Source: Endline learner survey.

Base: n = 185 (54 treatment and 131 control).

Some shorthand used in this figure. [c] indicates that data has been suppressed to protect anonymity.

These findings were supported in interviews with treatment learners. Many treatment learners interviewed were impressed by the quality of teaching and teaching approaches used during their FSQ Level 1 maths course. They appreciated how their tutors accommodated visual, auditory and tactile learners by demonstrating methods in multiple forms and providing more in-depth explanations. Some treatment learners also described their tutors prioritising learners' understanding on an individual level, often going over the same topic or method multiple times until everyone felt comfortable with their understanding. If a learner still struggled with a topic, the tutor would sometimes spend extra time with them after class and assign extra homework to target the specific method. Many treatment learners interviewed also enjoyed being assigned group work and supporting one another during the sessions as this helped them understand the content better and form friendships with like-minded individuals.

Overall, the treatment learners interviewed reported enjoying the course, stating that it exceeded their expectations. They found the volume of the content appropriate, often noting that the work was enough to help them understand the methods while being a manageable amount.

Learners surveyed generally found the level of maths covered in their course appropriate. This was true for both treatment (83% rated the level of maths as "about right") and control (79% rated the level of maths as "about right") learners (Table 23). The differences between groups were not statistically significant.

Table 23: Would you say that the maths covered on your course was...

Response option	Count – treatment	Percentage – treatment	Count – control	Percentage – control
Too easy	[c]	[c]	16	[u]
About right	45	83%	104	79%
Too difficult	[c]	[c]	8	[u]
Don't know / can't remember	[c]	[c]	[c]	[c]

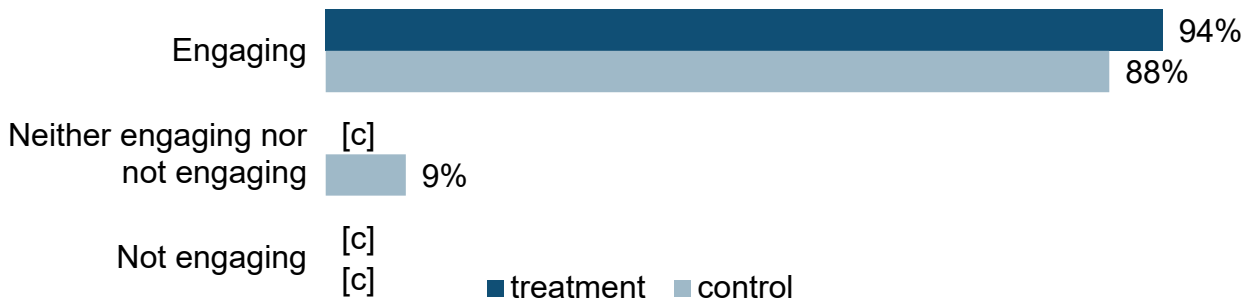
Source: Endline learner survey.

Base: n = 185 (54 treatment and 131 control).

Some shorthand has been used in this table. [u] indicates that data are not reported because the unweighted base is fewer than 50 and [c] indicates that data has been suppressed to protect anonymity.

Learners surveyed generally found their course engaging, which was true for both treatment (94%) and control learners (88%) (Figure 8). This difference was not statistically significant.

Figure 8. Overall, how engaging did you find the course?



Source: Endline learner survey.

Base: n = 183 (54 treatment and 129 control).

Some shorthand used in this figure. [c] indicates that data has been suppressed to protect anonymity.

In interviews, most treatment and control learners also reported finding most of the sessions engaging. Learners interviewed generally reported that sessions about topics they understood were more engaging as they were able to follow along and enjoyed solving the course work.

This strong satisfaction with the course was reflected in the high proportion of learners surveyed who reported that they would recommend it. A substantive percentage of learners had already recommended the course to people they knew (31% in treatment and 25% in control) or were likely to recommend it (55% in treatment and 67% in control). These differences were not statistically significant.

Stakeholder perceptions and engagement with the intervention

Many providers who were interviewed reported signing up for the trial because of the potential for access to free, high-quality CPD for tutors. They spoke of being keen to develop their tutors' practice and engage with new approaches that could improve outcomes for learners.

“Most tutors take responsibility for finding CPD and developing themselves, but there will be tutors that just tick along and don’t access anything. So, I think if there is national stuff that could be accessed, that you know is good quality, then we can recommend that to our tutors.” - *Provider, control group*

Among the treatment group, many providers planned to continue applying the AMA approach, suggesting that it meets needs within the FSQ Level 1 maths course. Some tutors and providers interviewed spoke of not only delivering the intervention with future cohorts of Level 1 FSQ maths learners but extending to other groups of learners as well. In line with the adaptations discussed earlier, some planned to continue delivering the

intervention with some “tweaking” to make it more applicable to their learners or delivery context. Many treatment providers interviewed also said they would recommend the approach to other adult learning providers. This suggests that the intervention was seen as feasible and useful within the context of the FSQ Level 1 maths course.

“We need to use this for Level 1, Level 2, and if we have a GCSE class in the future. Because I just think it’s such an amazing course ... The learners love it.” - *Tutor, treatment group*

3.7. Readiness for trial

This section explores the findings around scalability and replicability that indicate whether the intervention is ready to be trialled at larger scale. It also highlights lessons learned on the methodology and data collection methods.

3.7.1. Intervention readiness

Though most providers and tutors interviewed reported positive experiences with the lesson resources and delivering the intervention, they suggested a wide range of improvements to address challenges arising during delivery and improve the overall intervention. Some tutors interviewed reported errors in the materials, including mismatches between handouts and slides, inconsistencies in questions and answers, and missing examinable content. They felt that the resources needed more practice questions, clearer introductions and terminology, additional exam-style content, and a review to ensure all FSQ Level 1 maths examinable topics were fully covered. Some also felt that more guidance within the lesson plans would improve preparation and ensure more consistent delivery, especially as tutors needed to understand why the approach worked rather than just following the steps.

“Sometimes the learners were like, 'What's this? What's this? What's this? This is different. This is different from what I did. That's wrong.' If you could just have someone to always check that out before they send it.” - *Tutor, treatment group*

Many tutors interviewed also highlighted the need to adapt the materials for a wider range of learner levels. Some recommended creating a Level 2 version of the course, while others emphasised the importance of including foundational content for learners entering Level 1 with gaps, particularly those moving up from Entry. Many tutors interviewed felt that the current programme assumes Level 1 learners already have the underpinning skills in place, which they noted is often not the case. Several tutors interviewed also asked for clearer delivery guidance, particularly explicit explanations within lesson plans about the purpose behind specific steps and activities, and how they fit within the AMA principles.

Though the intervention was designed to be delivered face-to-face, providers were permitted to use AMA online if that was their usual course delivery style. However, online delivery was a major challenge, with many tutors who delivered the intervention online having to substantially adapt the resources for virtual teaching, as many activities were designed for in-person classroom use. Some spent significant time preparing slides, adjusting activities, and adding extra explanations to make online delivery workable. Others noted that certain practical or hands-on elements were easier to deliver face-to-face, where physical resources and real-time checking of learner understanding were more feasible. Therefore, either the intervention should be designated as suitable only for face-to-face classes, or materials for online learning should be made available.

3.8. Perceived outcomes

This section describes the evidence on perceived outcomes for tutors and learners, based on the perceptions of tutors and learners themselves. This section includes consideration of outcomes that are not examined in the impact evaluation.

3.8.1. Perceived impact on tutors

Tutors interviewed generally reported positive experiences delivering the intervention and identified several ways it influenced their teaching practice. Many adopted specific methods, most notably ratio tables, bar models, and chunking. Several tutors found these approaches helped learners make sense of topics such as percentages and division. Some tutors interviewed, while initially questioning the sequencing of topics, said the intervention prompted them to reconsider the order in which they introduced concepts. Some reported already using these techniques in other courses, while others said they plan to use them in future delivery.

“This ratio table, this is quite effective. I started using it for my other courses as well. So, this is quite useful, and how it linked with other topics in maths. That's brilliant.” - *Tutor, treatment group*

In the interviews, some tutors also noted that the structured PowerPoints and ready-made resources reduced preparation time. Several said the materials prompted them to make lessons more interactive, incorporate practical and group activities more often, or try new ways of presenting familiar concepts.

Some tutors interviewed also reported a gradual shift in their pedagogy and mindset as they became more familiar with the approach. While some of these tutors expressed caution at first, many said their confidence grew as they saw learners respond positively and make clearer connections between topics. The approach reinforced their belief that learners need to understand why methods work, not just how to apply them. They noted that the approaches made topics more accessible to adults who lacked confidence or had been out of education for many years. Several tutors interviewed said the experience

refreshed their practice and made lessons more varied and engaging. Although some tutors interviewed noted that time pressures limited how deeply they could explore concepts, many felt the intervention had a positive influence on their teaching across different levels.

Overall, tutors interviewed felt positive, motivated, and professionally strengthened by engaging with the intervention. Many valued the opportunity to learn new techniques, share practice, and reflect on their teaching, with some noting strong team interest and engagement as the trial progressed.

3.8.2. Perceived impact on learners

This section outlines the perceived impacts on learners based on data from both learners and tutors. The first section sets out the results from the learner surveys, both treatment and control, while subsequent sections describe the findings from the learner interviews, separated into perceived impacts on treatment learners and perceived impacts on control learners.

Survey results

In the endline survey, learners in both the treatment and control groups reported high levels of confidence working with numbers in everyday life. Table 24 shows that 82% of treatment learners at endline felt confident, compared to 88% of control learners. This difference was not statistically significant.

Table 24: Overall, how confident do you feel working with numbers in everyday life?

Response option	Count – treatment	Percentage – treatment	Count – control	Percentage – control
Very confident / Fairly confident	48	82%	128	88%
Neither confident nor not confident	5	[u]	11	[u]
Not very confident / Not at all confident	6	[u]	5	[u]
Not relevant to me	0	0%	[c]	[c]

Source: Endline learner survey.
Base: n = 204 (59 treatment and 145 control).

Some shorthand has been used in this table. [u] indicates that data are not reported because the unweighted base is fewer than 30 and [c] indicates that data has been suppressed to protect anonymity.

Table 25 shows that levels of confidence did not vary substantially between the control and treatment groups across a range of everyday numeracy tasks, such as identifying the best deals when shopping (86% confident in treatment group vs 90% confident in control group) and managing personal finances such as keeping track of bank account balances (91% vs. 93%). None of these differences were statistically significant.

Table 25: How confident do you feel about doing the following things in everyday life?

Response option	Count – treatment	Percentage – treatment	Count – control	Percentage – control
a) Checking your change is right when you have bought something	53	91%	131	92%
b) Working out the best deals when shopping	50	86%	129	90%
c) Helping children with maths homework or talking about maths / numbers with children	26	[u]	55	77%
d) Understanding interest rates on bank statements	44	76%	103	72%
e) Keeping track of your bank account balance	53	91%	133	93%
f) Working with numbers as part of a job	29	[u]	67	85%

Source: Endline learner survey.

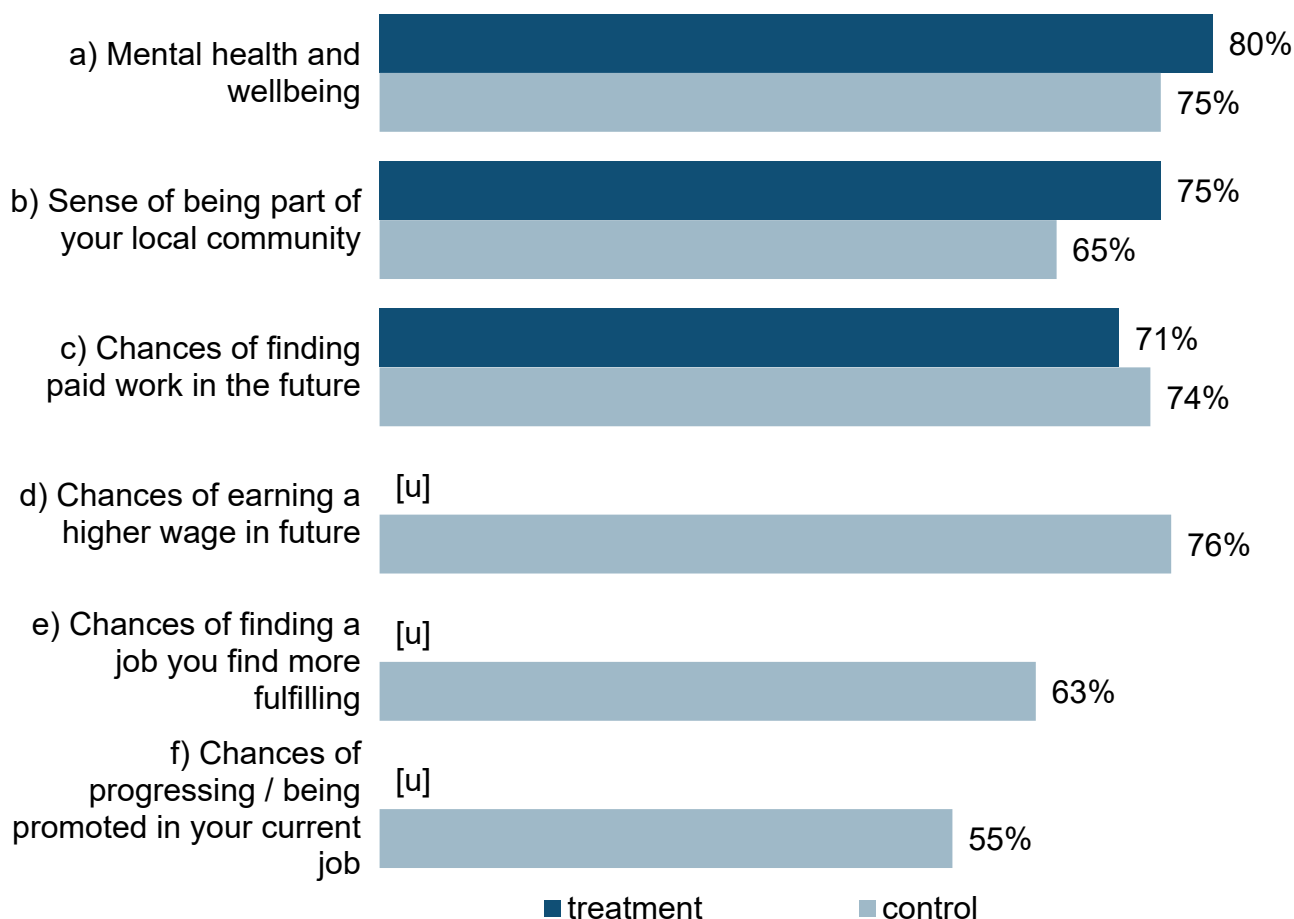
Base: questions a, b, e – n = 201 (58 treatment and 143 control), question c – n = 102 (31 treatment and 71 control), question d – n = 202 (58 treatment and 144 control), question f – n = 113 (34 treatment and 79 control).

Some shorthand has been used in this table. [u] indicates that data are not reported because the unweighted base is fewer than 30.

Note: Results show what percentage of everyone who answered the question gave 1 of the positive response options ("Very confident", "Fairly confident").

Figure 9 shows that both control and treatment learners surveyed at endline reported positive impacts from participating in FSQ Level 1 maths on a range of outcomes including wellbeing, sense of community, and future job prospects. There were no statistically significant differences between the 2 groups on these measures.

Figure 9: What impact, if any, do you think the course has had on your...



Source: Endline learner survey.

Base: questions a-c – n = 194 (56 treatment and 138 control), questions d-f – n = 108 (33 treatment and 75 control).

Some shorthand has been used in this figure. [u] indicates that data are not reported because the unweighted base is fewer than 30.

Note: Results show what percentage of everyone who answered the question gave 1 of the positive response options ("Very positive", "Fairly positive").

Table 26 shows that learners surveyed at endline also reported positive impacts across a range of maths outcomes related to AMA. However, these impacts were reported equally by both treatment and control learners. There were no statistically significant differences in self-reported maths outcomes between the 2 groups.

Table 26: What impact, if any, do you think the course has had on your...

Response option	Count – treatment	Percentage – treatment	Count – control	Percentage – control
Understanding the connections between different maths topics	53	95%	122	88%
Feeling willing to attempt maths problems	51	91%	124	90%
Having a deep understanding of maths concepts	50	89%	126	91%

Source: Endline learner survey.

Base: n = 194 (56 treatment and 138 control)

Note: Results show what percentage of everyone who answered the question gave 1 of the positive response options ("Helped a lot", "Helped a little").

Table 27 shows that most learners planned to do more study or training. This was similar for treatment (78%) and control learners (85%). Figure 10 shows that many learners said their experience of the course had influenced their decision to do more study or training, with 69% in the treatment group and 68% in the control group reporting they were influenced “a lot”.

Table 27: Are you planning to do any more courses or training?

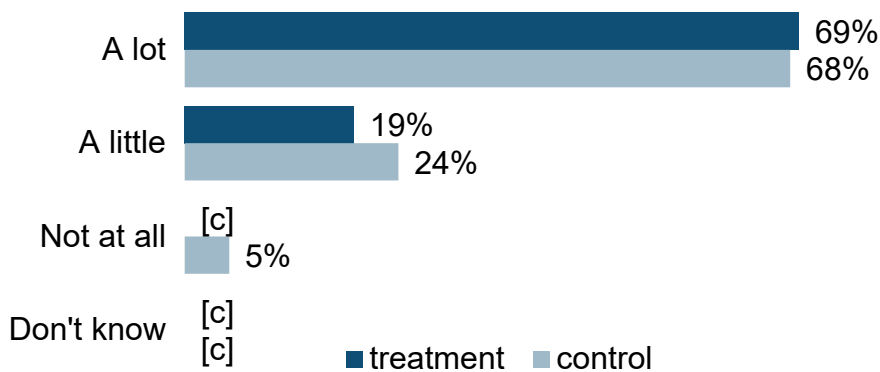
Response option	Count – treatment	Percentage – treatment	Count – control	Percentage – control
Yes – I plan to do more study or training	42	78%	112	85%
No	7	[u]	12	[u]
Don't know	5	[u]	6	[u]

Source: Endline learner survey.

Base: n = 185 (54 treatment and 131 control).

Some shorthand has been used in this table. [u] indicates that data are not reported because the unweighted base is fewer than 30.

Figure 10. To what extent, if any, did your experience of the course influence your decision to do more study or training?



Source: Endline learner survey.

Base: n = 154 (42 treatment and 112 control).

Some shorthand used in this figure. [c] indicates that data has been suppressed to protect anonymity.

Interview findings: treatment group

Confidence and attitudes

Interviews with tutors and learners suggested substantial gains in treatment learners' confidence with maths over the course of the programme. Many learners interviewed started the course anxious, reluctant to participate, or convinced they could not do maths. By the end of the course, many tutors interviewed reported that learners were more comfortable attempting questions, asking for clarification, and challenging answers, including from their tutors.

"They're challenging me for the answer. Why have I done this? Why have I given them the answer this way when I've worked it out that way?" - *Tutor, treatment group*

Some learners interviewed shared that tutors helped them develop their confidence by creating a positive and supportive learning environment. They felt that their tutors' confidence in them and reassuring approach helped them believe they could succeed in completing and passing the course. This helped build more positive attitudes towards maths.

"I don't know if it was down to the teaching or the content. Obviously, you're in a lot smaller class, and I think it was a lot more of a positive experience. I think it came across as being quite doable, we would start on easier and then gradually sort of gain in difficulty with things, so, I think that was really good for me...I think we were given that time and teaching to sort of develop our skills, feel confident, and obviously I think that's probably something that I just needed to do, really, through maths, we were sort of gaining confidence with it. I can do it, it's just believing I can." - *Learner, treatment group*

Several tutors interviewed also highlighted cases where learners who had previously avoided maths were now arriving at lessons more relaxed and willing to try. Some described transformations from learners entering the classroom “shaking and in tears” to learners who felt ready for exams and talked confidently about progressing to the next level. Many learners interviewed felt their confidence improved partly because of improved maths outcomes, but also because their knowledge improved. They reported that, as the sessions progressed, they gained a better understanding of the topics. This allowed them to believe in themselves and their maths abilities as they had tangible proof of their progress.

Many tutors and learners interviewed also reported that learners became more willing to attempt a wider range of questions and engage in problem-solving. Several tutors interviewed noted that learners asked more questions, tried different strategies, and used tools such as ratio tables or diagrams without prompting. They also described learners becoming more independent by checking their own work, identifying which maths skills they needed for a task, and choosing appropriate strategies with greater confidence. Some learners interviewed reported more willingness to attempt maths problems themselves rather than relying on their tutor or the internet, while others showed greater comfort exploring approaches even if they risked getting the answer wrong. Several tutors who participated in interviews highlighted that lower-level learners required more explanation and practice, but overall felt the approach encouraged greater willingness to tackle unfamiliar problems.

Adoption of AMA principles and understanding

The intervention was also seen to strengthen learners’ understanding of mathematical structures by helping them recognise clearer links between topics. Some tutors interviewed described learners realising that fractions, decimals, percentages, probability, and proportions rely on the same underlying ideas, with visual tools such as ratio tables supporting this shift in mindset. They reported learners became more confident identifying links between lessons, applying the same strategies across topics, and recognising patterns that previously seemed unrelated. Some tutors interviewed also described learners becoming better able to break problems down and check their work. While some lower-level learners required more time or support, tutors who participated in interviews generally felt that the approach promoted stronger conceptual understanding than relying on memorised formulas alone.

“I think it has given them a deeper understanding... they've said things like, 'Oh I knew that's how you did it, but I didn't understand why.' And now they see the why, so I think then, that will help them with taking it from one topic to another topic.” -
Tutor, treatment group

Skills for employment and life tasks

Many tutors interviewed reported clear improvements in learners' ability to apply maths in everyday and work-related situations. Learners were described as using percentage and proportional reasoning while shopping, comparing prices, and breaking problems down using ideas drawn from ratio tables. Several tutors interviewed noted that learners reported using the methods with their children at home, finding the visual or step-by-step approaches helped them understand more clearly than memorising formulas. Others gave workplace examples, such as a learner who worked for the NHS recognising connections between class content and tasks like calculating Body Mass Index (BMI).

Many learners interviewed felt that the course had helped them develop skills applicable to their careers and everyday lives. For learners who had joined the course to achieve a professional qualification and change careers in future, many expressed how crucial the FSQ Level 1 maths course had been for their professional development. Learners' success on the course combined with their tutor's belief in them inspired some learners to pursue other courses in their career field. For example, a learner shared that the knowledge and confidence gained through the FSQ Level 1 maths course had enabled them to pursue a culinary course.

"It gave me a bit more, obviously, confidence now, because I wasn't really sure about a lot of maths stuff. But it has given me a bit more confidence to do things now. Obviously, I'm doing a cheffing course, sometimes maths does become involved when you're weighing out ingredients, so yes, that has benefited me that way." - *Learner, treatment group*

Some other learners interviewed described how improved maths skills helped strengthen their work performance, while those who were retired or had jobs that did not require maths explained how the course helped them with everyday tasks such as budgeting or helping their children with homework.

Interest in taking up further maths courses

Learners' interest in progressing to further maths courses appeared to increase over the course of the intervention. Some tutors interviewed described learners who initially lacked confidence or explicitly stated they "weren't maths people" later expressing clear intentions to move on to Level 2 or GCSE courses. Several tutors interviewed noted that learners felt they had a more secure foundation and were motivated by both their progress and improved confidence. For some learners, especially those seeking qualifications for work, tutors interviewed felt the course strengthened their desire to continue to higher levels. Some tutors interviewed also described cases where previously anxious or disengaged learners gained enough confidence to register for Level 2 immediately after completing Level 1. Overall, interviews with tutors suggest that the intervention supported a stronger appetite for progression among many learners.

“And they're confident and they want to move on. They're talking about, 'When I progress to the next level, when I do this, when I do that.' Because they know they can do it. And they feel that they've got a really solid groundwork to progress.” - *Tutor, treatment group*

This was also reflected in interviews with learners. Many shared that the course made them interested in pursuing further maths courses, including Level 2. Many reported no longer fearing maths and that their newfound confidence had motivated them to continue with maths courses. Some learners were interviewed after passing Level 1, which reinforced this increase in confidence, and they had already started their Level 2 course. Some of those still completing the Level 1 course expressed excitement about potentially passing so they could move on to Level 2.

Unanticipated outcomes

The intervention seemed to provide some learners with a “lightbulb” moment, which tutors felt was an unanticipated outcome. In these “eureka” moments, learners who had struggled for years suddenly grasped a concept or achieved unexpectedly high mock scores. No negative unintended outcomes were reported. Tutors were generally unaware of any negative impacts, and none were identified during the research.

Variation by learner characteristics

Some tutors reported variation in perceived outcomes depending on learners' individual characteristics, including prior attainment, gaps in foundational skills, SEND, and language background. For learners with SEND, AMA was seen to be especially helpful for those who benefited from visual or structured methods. Some tutors who participated in interviews reported that the approach was sometimes helpful for ESOL learners when visual representations reduced the language load, though some noted that word problems continued to pose inherent language challenges regardless of the method used. Some tutors interviewed also noted that learners with very limited prior knowledge or large gaps in education sometimes struggled to absorb the volume of new information, making the methods less suitable without additional foundational work.

“It's the basic assumption that when somebody comes in at Level 1, they've got all that underpinning in place, and it's not always the case, is it? So, a good grounding in the basics to start with may be useful, because I think this course assumes that those things are in place, and it's not always so.” - *Tutor, treatment group*

Interview findings: control group

Most tutors interviewed from the control group reported a range of positive outcomes for learners from participation in the BAU FSQ Level 1 maths course. They described steady improvements in learners' confidence, understanding and willingness to attempt maths tasks, noting that adult learners were generally motivated, engaged and supportive of

one another. Many tutors said learners became more comfortable contributing in class, working in small groups and attempting problems independently. While some learners continued to find certain topics (such as fractions) difficult to apply in real-life situations, most were seen to build confidence as they realised that they could succeed in a subject many had previously found intimidating.

“I think they've done them good, because they're doing something very often they thought they couldn't do, it gives them confidence, they're building up the skills and seeing that they could do stuff that they perhaps previously thought they couldn't do. It's a great confidence builder.” - *Tutor, control group*

Some control tutors interviewed also highlighted wider benefits that extended beyond the classroom. Learners were described as applying maths more effectively in everyday life, including managing money, shopping, and measurements, and several were observed progressing into employment or further study. Findings from control learner interviews supported these views, especially regarding increased confidence and the ability to apply maths more effectively in everyday life.

ESOL learners were noted to progress successfully through multiple levels over time, with many tutors interviewed reporting that improved maths skills supported both day-to-day problem-solving and broader UK life. Some ESOL learners interviewed from the control group felt less anxious about applying for jobs after completing the FSQ Level 1 maths course as they felt better qualified for the positions they were applying for. Some tutors interviewed also described specific examples of ESOL learners using maths at work or reporting that the course had helped them handle practical tasks more confidently. Employed ESOL learners from the control group shared specific examples of how the calculations learned on their course had helped them understand and perform their roles better.

4. Conclusions

Before discussing the results further, it should be noted that this pilot trial was one of the first of its kind in the adult education sector. It has provided valuable lessons on both AMA specifically and running trials in the adult education sector.

The impact evaluation did not find any statistically significant results for the primary outcome at the 95% confidence level, which is the conventional threshold used in research to judge whether a finding is likely to reflect a real effect rather than chance. The results of the CACE analysis for the primary outcome showed a small positive increase in the estimated percentage point impact and a slight improvement in the significance of these effects. However, the CACE results were not statistically significant. It is important to note that any conclusions are limited by the small effective sample sizes.

The results for the secondary analysis were inconclusive. A positive effect was found for the composite measure for confidence in maths, though the effect for the main measure was negative. Neither result was statistically significant at the 95% confidence level, though the impact on the composite measure was statistically significant at the 90% confidence level. In practical terms, this means that the evidence was not strong enough to meet the usual 95% confidence threshold, but using a less strict 90% confidence level, the measure showed some evidence of effect. This suggests a possible impact, but the evidence is relatively weak and should be interpreted with caution. Overall, both maths confidence analyses were limited by the substantially lower sample size due to attrition, so these results are presented with low confidence.

At the same time, the IPE found there was a high level of compliance among tutors with the training requirements and delivery of intervention sessions. Among those with available monitoring data, most tutors (83%) attended all 4 training sessions while the remainder attended 2 or 3 sessions out of 4, and almost all tutors (95%) reported delivering each of the 15 sessions to learners. 69% of tutors attended at least 8 out of 12 lesson study sessions. A total of 233 learners (50%) in the treatment group attended 12 or more sessions out of 15, i.e., more than 80% of sessions, which was the threshold for compliance in the impact analysis.

However, there was also a high level of attrition in the trial, both in the treatment and control group. In the treatment group, 6 out of 19 providers withdrew after randomisation (32% attrition rate), and in the control group, 7 out of 19 providers withdrew after randomisation (37% attrition rate). Reasons for withdrawing among treatment providers included no longer having the required resources to engage in the intervention and no longer feeling the intervention aligned with their current delivery model. Among control providers, most simply did not engage in any trial activities and withdrew due to non-compliance, while 2 providers could not recruit enough learners to deliver the course. The high level of attrition suggests recruitment mechanisms should be improved in any future research.

Interviews with treatment tutors suggested that they largely followed the intervention lesson plans, though many needed to make small adaptations to account for learners' different needs, mode of delivery, and gaps in coverage of the curriculum. Most treatment tutors felt there were many elements of the intervention that were different to their BAU teaching approaches. Similarly, most control tutors were not aware of mastery principles, but some key features of the approach were present when they described their teaching style. Overall, the evidence suggests that the intervention does differ from BAU teaching approaches, and contamination is unlikely to have been an issue.

Most treatment tutors interviewed also reported positive outcomes from their experiences of delivering the intervention, including improved teaching practices, shifts in their pedagogy and mindset, and professional development. Similarly, many tutors felt learners had gained confidence during the course, with some more willing to attempt maths and problem solving. Many treatment tutors also observed evidence of learners benefiting from and adopting AMA principles, such as strengthening their understanding of mathematical structures and recognising clearer links between topics. A perceived increase in confidence was also reported prominently by learners interviewed and surveyed. It was reported, by many tutors and learners, that the course had positive impacts on learners' ability to apply maths in everyday situations and careers, and interest in progressing to further maths courses. Some tutors reported some variation in perceived outcomes and saw the intervention as particularly supportive for learners with SEND, as they benefited from visual and structured methods. However, it should be noted that improvements in outcomes were also reported prominently among control group tutors and learners, including in surveys with control group learners.

Overall, the findings suggest the intervention is practical and feasible to deliver, based on insights from those who completed the programme. However, some barriers and potential improvements were identified. Treatment tutors and providers were overall supportive of the approach, with many planning to continue or even expand delivery of AMA. Treatment tutors were generally satisfied with the lesson materials and training, and there was high satisfaction among treatment learners interviewed, including relating to the quality of teaching and the level of difficulty. However, many treatment tutors reported needing to make minor adaptations to account for learners' different needs, mode of delivery, and missed areas of the curriculum, while a small number of tutors made more significant adaptations. Overall, treatment learners reported enjoying the course, sharing that it exceeded their expectations. They found the volume of the content appropriate, often sharing that the work was enough to help them understand the methods while being a manageable amount.

It must also be noted that there are substantial limitations with the study, which require careful interpretation of the findings. For the impact evaluation, it is important to note that this was a pilot trial, with a small number of participants, so the trial was not sufficiently powered to detect small to moderate effects. This means that modest but potentially meaningful benefits from the intervention were unlikely to have been detectable and it is not definitive proof of no impact. The final sample size used in the analysis was substantially lower than the target number of providers and learners, reducing the

statistical precision of the results. In addition, outcome data were not available for all participating learners, which likely attenuated the estimated effects. As such, the trial was exploratory in nature and not statistically powered to detect treatment effects. Additionally, the impact analysis excluded 2 secondary outcomes (maths skill and attendance) due to data collection issues.

For the IPE, changes were made to the final design compared with the trial protocol. These include removing the endline survey of tutors to reduce practical and burden concerns, removing the observation of tutor training sessions as the delivery organisation feared it would affect participation, and removing the test of maths skills due to low uptake of the optional assessment and issues related to timing. Furthermore, the IPE had other limitations related to interview sampling, learner survey response rates, and the quality of management information. Therefore, while the IPE findings provide valuable insights into the range and diversity of views and experiences among different participants, they may not represent the wider population of learners and tutors taking part in the trial.

4.1. Lessons learned for future delivery

Though the intervention was generally well received, several challenges were identified that may help explain why these positive experiences did not translate into measurable impacts in the quantitative analysis. Some treatment tutors and providers interviewed noted that further support and development could be provided to improve curriculum alignment, applicability for learners at different levels, and online delivery. The trial also experienced high attrition among both treatment and control providers after randomisation. This not only reduced statistical power to detect smaller impacts, but also raised concerns about the practical burden of the intervention, particularly the time and resource demands associated with the tutor training. Some treatment tutors and providers reported that the mandatory lesson study sessions were overly burdensome and more flexibility in the training requirements would make it easier for providers and tutors to engage. These findings suggest that the training burden of the intervention should be reduced if implemented in future.

4.2. Considerations for future research

While the impact evaluation did not find statistically significant impacts, the pilot trial provides helpful evidence about the feasibility of running a future RCT on mastery approaches in adult learning settings. The study demonstrates that providers and learners can be successfully recruited, that randomisation at the provider level is workable, and that key outcome data can be collected. However, for potential future research, adaptations should be made to the intervention to improve retention and fidelity, including reducing the training burden and making the course materials more appropriate to a range of delivery models, such as online delivery. Any future trial design must also ensure a sufficient sample size for adequate statistical power to detect

treatment effects. Because the impact in this pilot trial was so small (an approximately 0.1 percentage point increase in the pass rate) it is not feasible to estimate the sample size required to detect this effect in any future trials. However, based on an assumed 10 percentage point difference between treatment and control (a conventional effect size estimation), a trial with the same patterns of variation and attrition rates would require approximately 241 providers across both arms, or 44% of all providers across England (out of the available 545). This is an extremely high number, and is unlikely to be a feasible recruitment target for a future trial. Further, the imperfect take-up rates prior to randomisation also supports the need to recruit an unfeasible number of providers if a future trial were to randomise at the provider level. Therefore, future research should consider other levels of randomisation (such as individual, class or tutor), while acknowledging the risks of contamination this could introduce.

This trial was part of a wider programme of Adult Numeracy Trials commissioned by the DfE. Alongside the individual trial reports, the DfE has published a programme evaluation report on findings related to running RCTs in the adult learning sector, describing the broader learnings for the sector (Mackay et al., 2026).

5. References

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

Carpenter, J. R., & Smink, J. (2020). Guidelines for reporting of multiple imputation. *Biometrical Journal*, 63(5), 874–878. <https://doi.org/10.1002/bimj.202000196>

Centres for Excellence in Maths (2023). *Mastery Teaching in Further Education: a handbook for practitioners*. Education and Training Foundation. Available at: https://etfoundation.co.uk/media/nl4i3toe/cfe_mastery_handbook_final.pdf (Accessed: 11 December 2025).

Department for Education (2025a). Further education and skills, Academic year 2024/25. Explore Education Statistics. <https://explore-education-statistics.service.gov.uk/find-statistics/further-education-and-skills/2024-25> (Accessed: 11 December 2025).

Department for Education (2025b). Further education participation, by ethnicity. GOV.UK Ethnicity Facts and Figures. Published 12 June 2025. <https://www.ethnicity-facts-figures.service.gov.uk/education-skills-and-training/a-levels-apprenticeships-further-education/further-education-participation/latest/> (Accessed: 11 December 2025).

Department for Education (2019). Improving attainment among disadvantaged students in the FE and adult learning sector: evidence review. <https://www.gov.uk/government/publications/improving-attainment-in-the-fe-and-adult-learning-sector/improving-attainment-among-disadvantaged-students-in-the-fe-and-adult-learning-sector-evidence-review-html#fn:3> (Accessed 13 March 2026)

Feeley N, Cossette S, Côté J, Héon M, Stremier R, Martorella G, Purden M. (2009). The importance of piloting an RCT intervention. *Can J Nurs Res*. Jun;41(2):85-99.

Guskey, T. R., and Pigott, T. D. (1988). Research on group-based mastery learning programs: A meta-analysis. *Journal of Educational Research*, 81, 197–216.

Imbens, G. W. & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*.

Jerrim, J., Austerberry, H., Crisan, C., Ingold, A., Morgan, C., Pratt, D., Smith, C., Wiggins, M. (2015). *Secondary Evaluation Report. Mathematics Mastery*. Education Endowment Foundation.

Jerrim, J., and Vignoles, A. (2015). *Mathematics Mastery: Overarching Summary Report*. Education Endowment Foundation.

Kulik, C. C., Kulik, J. A. and Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, 60, 265–299.

National Council for Further Education (2023). Functional Skills pass rates 2022-2023. Available at: <https://www.ncfe.org.uk/customer-and-learner-support/delivery-support/functional-skills/functional-skills-pass-rates-22-23/> (Accessed 30 January 2026).

Ritchie, J., Lewis, J. (2033) *Qualitative Research Practice: A Guide for Social Science Students and Researchers*, Sage Publications.

Vignoles, A., Jerrim, J., & Cowan, R. (2015). Primary Evaluation Report. Mathematics Mastery. Education Endowment Foundation.

Wake, G., Adkins, M., Dalby, D., Hall, J., Joubert, M., Lee, G., & Noyes, A. (2023). Teaching for Mastery Randomised Controlled Trial: Evaluation Report. Centres for Excellence in Maths. University of Nottingham.

White I, Royston P, Wood A. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*;30:377–399

Appendix 1: Glossary

Absolute standardised differences: A measure used to check if the treatment and control groups had similar characteristics at the start of a study. It compares the difference between the groups in a standardised way. A result below 0.1 (or 10%) usually shows the groups were well-balanced.

Adapted Mastery Approach for FSQ Level 1 in Maths: An intervention designed to support teachers of FSQ Level 1 Maths to adopt a 'Mastery' approach. This involves deepening understanding and spending time building from context to abstract reasoning, instead of just 'covering the curriculum'. It includes 15 taught sessions.

Adult Education Budget (AEB): The AEB provides government funding for skills training for adults aged 19 and over in England. It covers basic English, maths and digital skills. In 2024, it was renamed the Adult Skills Fund.

Adult education provider: An organisation that offers learning, training, and courses to adults aged 19 and over. These can include a wide range of institutions, such as Further Education (FE) colleges, local authority services, independent training providers (ITPs), and voluntary or community organisations.

Analysis sample: The final group of participants whose data is used to generate the final statistical results. It is often smaller than the initial sample due to exclusions (e.g. attrition, missing data, loss to follow-up).

Attrition rate: The percentage of participants or records lost between 2 points in a study (e.g. from baseline to endline).

Attrition: When participants leave a study before it is finished. This reduces the final sample size for analysis. This can occur when providers or learners withdraw, fail to complete assessments at the end of an intervention or submit required data. Attrition by trial arm looks at the attrition in the treatment and control groups separately. It is used to check for differential attrition (unequal loss between arms) that could bias comparisons.

Base size (base): The number of respondents that a particular statistic (e.g. a percentage or mean) is calculated from. In surveys, this is the number of responses.

Baseline: Data collected from participants at the start of a study, prior to any intervention. It serves as a benchmark for measuring change over time, checks that groups are balanced, and improves the precision of the final analysis.

Between cluster variance: The measure of variation in outcomes that occurs between different groups (or 'clusters') in a study (e.g. differences between providers). Larger values mean the clusters are more distinct from one another.

Bias: A systematic error that can make research results misleading. Unlike random chance, this error does not disappear by simply increasing the number of participants in the study.

Binary variable: A type of variable that can only have two possible values, such as 'pass'/'fail' or 'yes'/'no', often coded as 1 or 0 in statistical analysis.

Blinded: The practice of withholding information about group assignment to prevent conscious or unconscious bias. This ensures that outcomes are not influenced by the placebo effect (participants changing behaviour because they know they are being treated) or observer bias (researchers interpreting data differently based on their expectations). When information is withheld, a study is described as being 'blind'.

Business as Usual (BAU): The standard set of conditions or practices that participants experience if they are not assigned to receive a new intervention. BAU does not mean doing nothing; it means continuing with the existing approaches (e.g. the standard curriculum) rather than the new approach.

Complier Average Causal Effect (CACE): is a statistical estimate of the effect of an intervention specifically among those who complied with their assigned treatment condition. In this case, learners who attended at least 5 sessions and did not switch providers. Unlike the Intention-to-Treat (ITT) estimate, which measures the average effect across all participants regardless of whether they engaged with the intervention, CACE focuses only on compliers, typically producing a larger effect estimate. It is calculated using Instrumental Variable (IV) regression. CACE is useful for understanding the potential impact of an intervention under full engagement, but the estimate comes with a statistical penalty: it is based on a smaller, self-selected group, meaning the analysis is less powered to detect significant effects than the typical ITT analysis.

Computer Assisted Telephone Interviewing (CATI): A data collection method in which a researcher conducts a survey over the telephone while using a computer-based system to guide the interview and record responses.

Cluster Randomised Controlled Trial (RCT): A research method where groups of people (or "clusters"), rather than individuals, are randomly assigned to different trial arms.

Clustering: How individuals are naturally grouped together, such as learners within the same class or provider. Because people within the same cluster often share similar characteristics or experiences, statistical analysis must account for this to avoid drawing incorrect conclusions.

Coefficient (Standard error): In a regression analysis, the coefficient is the number that represents the size and direction of the relationship between a predictor variable and the outcome. The standard error is a measure of the statistical accuracy of that coefficient; a smaller standard error means a more precise estimate.

Cohen's d: A widely used statistical measure for calculating effect size, which shows the size of the difference between the averages (means) of two groups. A smaller number (e.g. around 0.2) indicates a small effect, meaning the difference between the two groups is minor. A larger number (e.g. 0.8 or higher) indicates a large effect, signifying a substantial and more meaningful difference between the groups.

Cohen's h: This is a standardised measure of effect size used to compare two proportions (i.e., the difference between two percentages). It is useful for comparing results across different studies that might have different sample sizes or baseline rates. A small Cohen's h (around 0.2) means a smaller difference between proportions. Even if the difference is statistically significant, it might not be very meaningful in a practical sense. A larger Cohen's h (around 0.8) means a more substantial difference between proportions.

Compliance: The extent to which providers adhered to the trial requirements outlined in the Trial Readiness Packs, including data submission requirements and adherence to eligibility criteria.

Confidence intervals: Confidence intervals are used to express the certainty of an estimate. The interval is the range of values within which the 'true' value for the whole group is highly likely to lie. The smaller the range, the more certain the estimate. A 95% confidence level (the most common) means that if we repeated the study and analysis 100 times, 95 of the ranges calculated would include the true value of the population.

Contamination: A potential issue in a trial where the control group is unintentionally exposed to the intervention.

Contextualised Curriculum for FSQ Level 1 in Maths: An intervention designed to improve adults' attitudes and beliefs about the value of maths and attainment in maths by using a Realistic Mathematics Education (RME, rme.org.uk) approach in classroom delivery. This means learners are engaged with context prior to learning a technique. It includes 12 taught sessions.

Continuing Professional Development (CPD): Ongoing learning activities for tutors/teachers to maintain and improve their professional knowledge and skills (for example, training, workshops, mentoring). Many adult education providers (e.g. FE colleges, training centres) require documented CPD for quality assurance and accreditation purposes.

Control group: A group that does not receive the intervention(s) being tested within an RCT design following randomisation. They are monitored alongside the group(s) receiving the intervention(s), and their results are compared to their treatment counterparts to understand what impact the intervention has had, compared to receiving no intervention. Any changes or effects detected within the control group over the course of the RCT can be interpreted as what would have happened normally.

Cooperation rate: A survey metric showing the proportion of contacted, eligible people who completed the survey. Typically calculated as completes divided by (completes + refusals + break-offs) among those reached and eligible.

Correlation: A correlation is a statistical measure that describes the strength and direction of an association between two or more variables. However, it is important to note that correlation does not imply causation, or how much such variables will change when a change is observed in the independent variable.

Covariates: Characteristics of participants (e.g. age, prior qualifications) that are measured at the beginning of a study and can be used in the statistical analysis to account for pre-existing differences between groups.

Descriptive statistics: Statistics used to summarise and describe the main features of a dataset, such as the mean, median, and standard deviation.

Dummy indicators: Also known as dummy variables, these are binary variables (using 0s and 1s) created to include categorical information (like which region a provider is in) in a regression model.

Effect sizes: A standard metric that quantifies the strength of a result. An effect size tells you not just if an intervention worked, but how much it worked, allowing for comparisons between different studies and contexts. Larger effect sizes indicate a stronger effect.

Embedding Maths in health and social care (Embedded Maths): An intervention to improve numeracy skills in adults by embedding mathematical concepts into the curriculum of Health and Social Care Level 2 qualifications.

Endline survey: A survey completed at the end of an intervention period.

Entry level: Qualifications at entry level provide an introduction to education and can lead to certification of essential skills and knowledge for beginners.

ESOL learners: 'English for Speakers of Other Languages' learners; those taking part in a course for whom English is not their first language.

Experimental: An evaluation design where participants are deliberately assigned to groups, ideally through randomisation. By ensuring that there are no systematic differences between the treatment and control groups at the start (such as control groups and treatment groups formed by individuals with very different ages), this design provides the strongest possible evidence that the intervention caused any observed results.

Exploratory analysis: An early analysis of data to find initial patterns or interesting results. Findings from this stage are treated as suggestions that need to be tested properly in a future study, as they have a higher risk of being due to chance.

Family Numeracy: A programme hosted by primary schools for parents and their children to improve adult numeracy skills through family learning. Includes 12 hours of learning: 6 in-person group sessions and home learning activities.

Feasibility: An assessment of whether a proposed study or intervention can be practically implemented given the available resources, time and logistical constraints. It focuses on practical considerations such as recruitment volumes and data collection processes.

Fidelity: The extent to which an intervention is delivered as intended. It assesses whether what was implemented aligns with the original design, ensuring that the results reflect the true intervention rather than a diluted or altered version.

Foundation Tier Maths GCSE: Maths GCSE entry for grades 1-5.

Functional Skills Qualification (FSQ) Level 1 in Maths: A qualification that focuses on practical mathematical skills needed for everyday life and work, equivalent to GCSE grade 1-3. FSQ Level 1 Maths is usually targeted at individuals who require a more applied or vocational approach to learning maths and is suitable for learners aged 16+ years in further education, apprenticeships, or adult learning.

Guided learning hours: The amount of time a tutor is scheduled to be present to provide specific guidance to learners as part of a course of study.

Hedges' g: Hedges' g is a standardised measure of effect size that expresses the difference between two means in terms of standard deviations. It includes a small correction for bias in small samples, making it slightly more accurate when the sample size is small. Around 0.2 is considered a smaller Hedge's g (a smaller difference between means), and 0.8 is considered a larger Hedge's g (a bigger difference between means).

Impact evaluation: A study designed to determine whether an intervention is the cause of an observed change. It works by comparing the outcomes of participants against a counterfactual – that is an estimate of what would have happened to those same participants had the intervention not taken place.

Impacts: The broader, long-term effects of an intervention on participants and their environment, such as improved employment prospects or sustained changes in teaching practices.

Implementation and Process Evaluation (IPE): A study designed to complement an impact evaluation by examining how the intervention was put into practice. While the impact evaluation measures outcomes, the IPE assesses factors like fidelity and participant engagement to explain why those results occurred.

Imputation: This is the statistical process of replacing missing data with substituted, plausible values.

Individualised Learner Record (ILR): The ILR is an on-going collection of data about learners from training providers in the Further Education (FE) and Skills sector in England. It was used in the Adult Numeracy Trials as a source of administrative data for outcomes such as grades, course completion and progression to further learning.

Instrumental Variable (IV) regression: A statistical method that estimates the effect of an intervention only on the people who actually took part in it. It is used to adjust the results when some people assigned to the treatment group did not participate.

Intention to treat (ITT): This is a fundamental principle for analysing the results of a randomised controlled trial. It means that all participants are analysed in the group to which they were originally randomised, regardless of whether they actually received the intervention, completed it, or complied with it. This method preserves the benefits of randomisation and avoids bias.

Interaction effects: When the effect of an intervention differs depending on another factor rather than the intervention itself (for example, the impact varies by delivery mode, provider, or learner characteristics). This is also referred to as moderation.

Intercept: In a regression model, the intercept is the predicted value of the outcome variable when all predictor variables are set to zero. It represents a baseline or starting point.

Intervention/ treatment arms: In a trial, the "arms" are the different groups to which participants are assigned. An "intervention arm" or "treatment arm" is any group that receives a specific treatment or intervention being tested.

Intervention: In the context of a trial, an intervention refers to a specific programme, approach, or set of activities being tested for its effectiveness. It represents the treatment or change being implemented with participants, which is then compared against a control group or alternative approach to measure its impact on specified outcomes.

Intraclass correlation coefficients: A 0–1 measure of how similar outcomes are within the same cluster (e.g. provider or class) compared with across clusters. Also known as intra cluster correlation coefficients.

Key stage: A term used in the education system in England, Wales, and Northern Ireland to describe a specific stage of compulsory schooling, defined by a child's age. Key Stage 1 covers ages 5-7. Key Stage 2 covers ages 7-11. Key Stage 3 covers ages 11-14. Key Stage 4 covers ages 14-16 and concludes with national exams (GCSEs). Key Stage 5 covers ages 16-18.

Kicktag: Kicktag is the brand name of the data collection system used for these trials, also known as the Ipsos Data Portal.

Learner progress: A measure of the change in learner outcomes over the duration of a trial. It can involve tracking of development in areas such as academic performance, confidence, and study skills amongst both the treatment and control groups, enabling comparison of the relative progress of each.

Level 1: Qualifications at level 1 include or are equivalent to GCSE grades 3, 2, or 1 (previously D, E, F or G).

Level 2: Qualifications at level 2 include or are equivalent to GCSE grades 9, 8, 7, 6, 5, or 4 (previously grade A*, A, B, or C).

Likert Scale: An ordered rating scale used to measure attitudes or self-reports (e.g. strongly disagree to strongly agree, or 1 to 6). Often used for confidence or satisfaction questions.

Linear regression: A statistical method used to estimate the relationship between an outcome variable (such as GCSE grade) and one or more predictor variables.

Local area: The collective term 'Local areas' covers the authorities that commission and coordinate Multiply programme delivery in their area. These are the Greater London Authority (GLA), Mayoral Strategic Authorities (MSAs) and upper tier and unitary local authorities outside of these areas.

Logistic link function: A mathematical function used in logistic regression to model the relationship between predictor variables and a binary outcome, ensuring the predicted value is a probability between 0 and 1.

Longitudinal: A type of research design that involves collecting data from the same subjects repeatedly over a period of time. This allows researchers to track changes, development, and long-term effects.

Lower bound/Upper bound: These are the lowest and highest values in a confidence interval.

Managed Service Supplier (MSS): The MSS led on the recruitment and management of providers and schools who took part in the course-based adult maths trials. They were responsible for contract management, monitoring and reporting of delivery and issuing payments. The MSS also facilitated tutor training for providers assigned to treatment groups.

Maths GCSE: An academic qualification typically taken by students in the UK around the age of 15-16, but which can also be taken by people of all ages. It is graded 1-9 with anything above grade 4 considered a pass. The GCSE serves as an important benchmark for further academic and professional pursuits, often required for entry into further education, vocational qualifications, or employment.

Mayoral Strategic Authorities (MSAs): Regional governance bodies in England that replaces the previous Mayoral Combined Authority model. Led by a directly elected mayor, these authorities coordinate wide-scale economic growth, infrastructure, and public services across a defined geographical area without replacing local councils. There was a transition from Mayoral Combined Authorities to Mayoral Strategic Authorities (MSA) as part of the 2025's Devolution Bill. This reflects a shift in UK devolution, moving from bespoke arrangements to a uniformed statutory framework.

Meta-analyses: Studies that systematically combine and statistically pool results from multiple studies on the same question to produce an overall estimate of impact.

Minimal detectable effect sizes (MDES): This is the smallest true effect (or impact) of an intervention that a study has a good chance (usually 80% probability) of detecting as statistically significant. It is calculated before a study is carried out (at the design stage) to determine whether the sample size is sufficient to find a meaningful result.

Missing at Random (MAR) pattern: A situation where data is missing in a systematic way, but the pattern behind that situation is explained by other variables in your dataset. For example, if men were less likely than women to answer a question about their confidence, the data would be considered MAR because the reason for the missingness (gender) is a known, recorded variable. Because we can see this pattern, we can use statistical techniques to account for it during the analysis.

Missing Not at Random (MNAR): A situation where a data point is missing specifically because of the value itself. For example, if people with lower confidence scores are less likely to answer a confidence-related question. This is the most difficult type of missing data to handle because it introduces bias that is hard to detect.

Mixed-methods design: A research approach that integrates quantitative metrics (such as survey responses) with qualitative inquiry (such as interviews) to validate findings. By triangulating different data sources, researchers can overcome the limitations of a single method to reach more robust, nuanced conclusions.

Multimode: This refers to surveys that can be completed through different channels or modes, e.g. online or by telephone.

Multiple Imputation by Chained Equations (MICE): A statistical method that handles missing data by creating several different plausible datasets. By analysing these combined datasets, researchers can account for the uncertainty of the missing values, leading to more accurate standard errors and conclusions.

Observational studies: Studies that observe what happens without assigning people to conditions (for example, tracking outcomes in naturally occurring groups). They are useful for describing patterns and associations but are weaker for evidencing causal claims.

Ordinal data: A type of categorical data where the categories have a natural, ordered relationship. For example, survey responses like 'No qualifications', 'Entry level', 'Level 1', 'Level 2', and 'Level 3 or above' are ordinal and follow that order.

Orthogonality: A statistical property in experimental design where the effects of different variables can be estimated independently of one another. A test of joint orthogonality is used to check if the randomisation process resulted in balanced groups.

Outcomes: The specific, measurable results of an intervention that are tracked to evaluate its effectiveness, such as learner pass rates, attendance, and changes in confidence.

Outputs: The immediate, tangible products or services delivered by an intervention as a direct result of its activities. They describe 'what was done' or 'what was produced' rather than the changes that resulted from it. For example, the number of learners who participated in a maths lesson.

Parameter: A numerical value that describes a characteristic of an entire population (e.g. the true average pass rate for all learners). In research, statistics from a sample are used to estimate these population parameters.

Pedagogy: The theory and practice of teaching and learning (for example, the methods, principles, and strategies used by tutors).

Percentage points (ppt): A percentage point is the unit for the absolute arithmetic difference of two percentages. For example, moving from 10% to 12% is an increase of two percentage points.

Pilot Randomised Controlled Trial (Pilot RCT): A small-scale randomised study to assess the practical application of an intervention and the validity of the research methods. It focuses on gathering evidence regarding feasibility, implementation, and acceptability, allowing researchers to refine the programme design and data collection tools based on real-world feedback.

Probit regression analysis: This is a statistical technique used to model binary (yes/no) or ordinal outcomes. It predicts the probability that an outcome will occur based on a set of predictor variables.

Purposive sampling: A sampling technique where researchers deliberately select participants based on specific characteristics relevant to the study's goals, rather than at random.

p-value: The p-value, or probability value, is the probability that a result occurred by chance. A small p-value (usually 0.05 or less) suggests the result is 'statistically significant', meaning it is unlikely to be a coincidence.

Qualitative data: Refers to non-numerical data that is descriptive in nature, such as interview transcripts, observations, and case studies. It focuses on understanding experiences and perspectives.

Quantitative data: Refers to numerical data that can be measured and statistically analysed, such as test scores, pass rates, and survey ratings.

Quasi-experimental: An evaluation design that attempts to estimate impact without using random assignment. Instead, it relies on statistical techniques (such as matching or difference-in-differences) to construct a comparison group that resembles the treatment group as closely as possible.

Randomisation: The process of assigning participants to treatment or control groups using a random mechanism (such as a computer algorithm). This ensures that every participant has an equal probability of assignment, creating groups that are statistically equivalent at the start of the study.

Randomised Controlled Trial (RCT): An evaluation design where participants are randomly assigned to either a treatment group or a control group. This process ensures the groups are statistically equivalent at the start, meaning that any difference in final outcomes can be confidently attributed to the intervention rather than external factors.

Readiness for trial: An assessment of whether an intervention is sufficiently well-developed and stable enough to meet the requirements of a randomised controlled trial, and whether the proposed research methods are feasible.

Regression model: A statistical tool used to model and analyse the relationship between a dependent variable (the outcome) and one or more independent variables (the predictors).

Robustness check: An analysis to test whether the main results of a study hold up when the methods or assumptions are slightly changed.

Robustness: Whether or not the main results of a study hold up when the methods or assumptions are slightly changed.

Sample frame: The list or source from which a sample of participants is drawn.

Semi-structured interview: A qualitative interview format that combines a pre-determined set of open-ended questions with the flexibility to explore new ideas. It ensures that key topics are covered for every participant, while still allowing the interviewer to probe deeper into interesting or unexpected answers.

SEND learners: Those taking part in a course who have Special Educational Needs and/or Disabilities.

Sensitivity analysis: A statistical method used to assess how the results of a study might change if key assumptions or population data were different.

Skewed: This describes a distribution of data that is not symmetrical. A distribution is "skewed" if the data points are not evenly distributed around the average (mean). For example, a sample may be skewed towards having lower levels of qualifications if the sample contains more data entries with lower levels than higher ones.

Standard deviation: This is a measure of the amount of variation or dispersion in a set of values. A low standard deviation means that all values tend to be close to the average (mean), while a high standard deviation means that the values are spread out over a wider range.

Statistically powered: A term describing a study that has a large enough sample size to have a high probability (typically 80% or more) of detecting a real effect if one exists.

Statistically significant: A result is "statistically significant" if it is unlikely to have occurred by random chance alone. Researchers typically set a threshold to make this determination (p-value). It suggests there is a real effect or relationship in the data.

Stratification: The process of dividing a population into distinct subgroups or "strata" (e.g. by region) before randomisation to ensure that each subgroup is appropriately represented in the treatment and control arms.

Stratified cluster randomisation: A research method that combines cluster randomisation with stratification to ensure balance among trial arms. Clusters (such as groups of people) are divided into distinct subgroups or 'strata' (for instance, region) before randomisation to ensure that each subgroup is appropriately represented in the treatment and control arms.

Technical Steering Group (TSG): A Technical Steering Group was established by DfE to provide advice, guidance and oversight of key design elements of the trials. They also had ownership of technical risks and oversight of final outputs and ethical considerations.

Theory of Change: A model that explains how an intervention is expected to work. It maps the logical pathway from the inputs and activities to the intended short-term outcomes and long-term impacts.

Three-armed parallel cluster randomised pilot trial: A research design with 3 groups that are tracked simultaneously ("parallel"). The randomisation happens at a group level ("cluster," e.g. a provider), and it is a smaller scale "pilot" study to test the feasibility of a larger trial.

TIDieR framework: TIDieR stands for the Template for Intervention Description and Replication. It is a 12-item checklist and guide designed to improve the completeness and quality of how research interventions are described in publications.

Treatment group: The group of participants randomly assigned to receive the specific programme or policy being tested. Their outcomes are compared against those of the control group to determine if the intervention caused a significant change.

Trial participant: An individual who meets the eligibility criteria, has provided informed consent, and has been formally enrolled or randomised into the study.

ULN: Unique Learner Number. A unique 10-digit number assigned to individuals over the age of 14 involved in education or training in the UK.

Variance: A specific statistical measurement that measures the spread of data points around their average value (the mean). A small variance means the data points are clustered tightly around the average, while a large variance indicates they are more widely scattered.

Vector: In a statistical context, a vector is an ordered list of numbers, often representing a set of characteristics for a single participant or the coefficients in a regression model.

Within cluster variance: The measure of difference in outcomes that occurs among individuals within the same group or "cluster" (e.g. among different learners at the same provider).

Appendix 2: Detailed intervention description

Initial training

Tutors attended 4 initial training sessions, each lasting 2.5 hours. These included 2 sessions on introducing the programme and mastery approaches and principles, and 2 sessions providing an in-depth look at the lessons and planning for delivery.

Lesson study sessions

Tutors attended weekly 2-hour sessions led by a trainer, giving tutors the opportunity to work together to reflect on teaching practices, deconstruct lesson content, plan for teaching of upcoming lessons, and consider the practicalities and possible challenges of delivering the lessons.

Lessons

Tutors delivered 15 sessions to learners, each focusing on a topic area. For each of these topics, tutors received lesson plans, slides, and handouts. Table 28 summarises the objectives of each lesson, and what parts of the FSQ Level 1 curriculum they cover.

Table 28: Overview of lessons

Name	Objective	Curriculum and activities
Session 1: Maths mindsets and resilience	<ul style="list-style-type: none"> • Get to know each other and recognise learners' previous maths experiences. • Categorise attitudes into fixed or growth mindsets. • Recognise that progress is maximised when in the growth zone. • Discuss strategies for reducing maths anxiety. • Use ratio tables to answer proportion questions. 	Lesson to be used at the start of the course to ensure that a positive learning environment is developed. <u>Purpose of activities</u> <ul style="list-style-type: none"> • Introduce the lesson. • Allow learners to share their previous experiences and understand that they are not alone. • Help learners understand their maths mindsets. • Explore myths around learning maths. • Explore the zones of the growth zone model and understand how it can be used as a tool to regulate emotions while learning maths. • Develop strategies to cope with anxiety or stress. • Introduce ratio tables as a tool to get 'unstuck'. • Review and recap.

Name	Objective	Curriculum and activities
Session 2: Graphs and charts	<ul style="list-style-type: none"> ● Record data in tally charts, frequency tables and grouped frequency tables. ● Present data as bar charts and line graphs. ● Interpret bar charts and line graphs. 	<p><u>Curriculum:</u> Handling information and data: 27 Represent discrete data in tables, diagrams and charts including pie charts, bar charts and line graphs. 28 Group discrete data and represent grouped data graphically.</p> <p><u>Purpose of activities</u></p> <ul style="list-style-type: none"> ● Introduce the topic of data collection and organisation. Tally charts and frequency tables are introduced. ● Organise data and create a tally chart and frequency table. ● Introduce more complex data which needs to be grouped. ● Ensure that learners understand the most appropriate method for different types of data as well as how to represent data using a line graph and a bar chart. ● Apply their knowledge to exam questions. ● Review lesson objectives and summarise key learning points.
Session 3: Averages and range	<ul style="list-style-type: none"> ● Understand how and when to calculate the mean. ● Understand how to calculate range and that range is the data spread and not an average. ● Check answers using reverse calculations. 	<p><u>Curriculum</u> Handling information and data 27 Represent discrete data in tables, diagrams and charts including pie charts, bar charts and line graphs. 28 Group discrete data and represent grouped data graphically.</p> <p><u>Purpose of activities</u></p> <ul style="list-style-type: none"> ● Introduce the topic of averages and range. ● Introduce how to calculate averages, range and reverse calculations. ● Calculate mean and range in context. ● Apply their knowledge to exam questions. ● Review lesson objectives and summarise key learning points.

Name	Objective	Curriculum and activities
Session 4: Direct proportion	<ul style="list-style-type: none"> Identify when 2 quantities differ in direct proportion to each other. Understand the multiplicative relationship between 2 quantities (non-calculator). Solve simple proportional problems using efficient methods with ratio tables. 	<p><u>Curriculum</u> 17 Work with simple ratio and direct proportion.</p> <p><u>Purpose of activities</u></p> <ul style="list-style-type: none"> Introduce the concept and meaning of direct proportion graphically. Formalise the use of ratio tables by exploring proportionate amounts of ingredients for cookie recipe. Explore problem-solving approaches using ratio tables. Procedural variation – practice the use of ratio tables using mastery approach to questioning. True or false – to draw upon common misconceptions. Find multipliers and dividers. Apply their knowledge to exam questions. Use the proportionality of conversion graphs to solve problems. Summarise learning to capture ways of thinking and to clarify the concept of proportional reasoning.
Session 5: Scales, maps and units	<ul style="list-style-type: none"> Convert metric units of measurement Use a scale to find lengths. Represent a proportional situation in a ratio table. 	<p><u>Curriculum</u> Using numbers and the number system: 17 Work with ratio and direct proportion. Using common measures, shape and space: 21 Recognise and make use of simple scales on maps and drawings.</p> <p><u>Purpose of activities</u></p> <ul style="list-style-type: none"> Introduce different metric units of length. Introduce the concept of measurement and how to change freely between different units using ratio tables. Misconception – not knowing when to divide or multiply before problem-solving tasks. Collaborative exploration to develop the use of scales given in the form 1: n to solve the dimensions of scale furniture. Understand how scale works on online maps. Attempt a selection of exam questions to practice exam technique. Summarise learning to capture ways of thinking and to clarify the concept of reading and interpreting scale, maps and units.

Name	Objective	Curriculum and activities
Session 6: Multiplication, division and estimation 1	<ul style="list-style-type: none"> • Explore, evaluate and select different representations for multiplication. • Apply various methods and representations to a singular context using integers. • Apply efficient mastery methods to questions in different contexts. • Apply estimation, inversion and rounding to check accuracy of answers. 	<p><u>Curriculum</u></p> <p>Using numbers and the number system:</p> <p>3 Multiply whole numbers and decimals by 10, 100, 1000.</p> <p>11 Multiply 2- and 3-digit whole numbers.</p> <p>12 Approximate by rounding to a whole number and one decimal place.</p> <p><u>Purpose of activities</u></p> <ul style="list-style-type: none"> • Hook: Discussion about why multiplication is needed. • Check learners understanding of place value positions. • Understand how to multiply and divide by 10 and 100 using the idea of place value. • Consolidate understanding of the effect of multiplying and dividing by a power of 10. • Knowledge identification / development of understanding behind the methods. • Consolidate learners understanding of how to round decimals to the nearest integer and to 1 decimal place. • Appreciate and select a method. • Consolidate learning. • Review lesson objectives and summarise key learning points.

Name	Objective	Curriculum and activities
<p>Session 7: Multiplication, division and estimation 2</p>	<ul style="list-style-type: none"> • Be able to multiply decimals. • Apply various division methods and representations to a context involving integers. • Apply inverse operations to division problems to check accuracy of answers. • Apply rounding to divisions in order to check accuracy of answers. 	<p><u>Curriculum</u> Using numbers and the number system: 3 Multiply and divide whole numbers and decimals by 10, 100, 1000. 4 Use multiplication facts and make connections with division facts. 11 Add, subtract, multiply and divide decimals up to 2 decimal places. 12 Approximate by rounding to a whole number or to 1 or 2 decimal places.</p> <p><u>Purpose of activities</u></p> <ul style="list-style-type: none"> • Recap what learners have learned about multiplication to address any remaining misconceptions. • Use knowledge of multiplying whole numbers and estimation to multiply decimals. • Assess prior knowledge and understanding. • Reinforce that division is the inverse of multiplication and introduce the use of the area diagram for division. • Ensure learners are able to check that their answers are sensible. • Apply the concepts learned to a real-life scenario. • Investigate whether a chosen method works for a 3-digit number. • Understand the meaning of a remainder in context. • Check understanding by spotting misconceptions. • Division with a decimal value. • Apply their knowledge to exam questions. • Review lesson objectives and summarise key learning points.

Name	Objective	Curriculum and activities
Session 8: Fractions of amounts	<ul style="list-style-type: none"> • Understand what a fraction is. • Find a fraction of an amount, with and without a calculator. • Understand how a bar model can be used as a tool for finding fractions of amounts. 	<p><u>Curriculum</u> Using numbers and the number system: 8 Read, write, order and compare common fractions and mixed numbers. 9 Find fractions of whole number quantities or measurements</p> <p><u>Purpose of activities</u></p> <ul style="list-style-type: none"> • Understand what a fraction is and to gauge learners' starting points. • Further understand what a fraction is by discussing parts of a fraction and clarifying possible misconceptions. • Use a bar model to find a fraction of an amount, without and then with a calculator. • Express amounts as fractions and use problem-solving strategies to solve a multi-step problem. • Check and consolidate understanding by answering exam questions. • Review lesson objectives and summarise key learning points.
Session 9: Sharing in a ratio	<ul style="list-style-type: none"> • Understand how ratios correspond with real-world situations. • Simplify ratios. • Use ratio reasoning to solve whole-to-part and part-to-part simple problems. • Use bar model representations to provide insights into solving problems. • Understand the importance of using reverse calculations to check answers. 	<p><u>Curriculum</u> 5 Check the sense, and reasonableness, of answers. 17 Work with simple ratio and direct proportion</p> <p><u>Purpose of activities</u></p> <ul style="list-style-type: none"> • Understand what a fraction is and to gauge learners' starting points. • Introduce the concept of ratio. • Assess prior learning. • Collaborative exploration and using chocolate chunks to develop the construction/use of bar models to write a ratio. • Develop and practice a method for simplifying ratios. • Move out of the concrete context of chocolate into the abstract context of a utility bill. • Deepen learners' understanding using bar models. • Check and consolidate understanding by answering exam questions. • Summarise learning to capture ways of thinking and to clarify the concept of proportional reasoning.

Name	Objective	Curriculum and activities
<p>Session 10: Fractions, decimals, percentages</p>	<ul style="list-style-type: none"> • Find equivalent fractions. • Convert between a decimal, fraction and percentage where the denominator is a factor of 10 or 100. • Understand and use equivalence between common fractions, decimals and percentages. • Order fractions, decimals and percentages by understanding place value. 	<p><u>Curriculum</u> Using numbers and the number system: 8 Read, write, order and compare common fractions and mixed numbers. 10 Read, write, order and compare decimals up to three decimal places. 13 Read, write, order and compare percentages in whole numbers. 16 Recognise and calculate equivalences between common fractions, percentages and decimals</p> <p><u>Purpose of activities</u></p> <ul style="list-style-type: none"> • Introduce the 100-grid representation and find equivalent fractions. • Convert between fractions and percentages. • Introduce the place value table as a representation to help understand place value of whole numbers and decimals. • Use the place value table to convert between decimals and fractions. • Understand place value in relation to decimals and then fractions and percentages. • Collaborative exploration to develop thinking and reasoning skills for fractions, decimals and percentages. • Check and consolidate their understanding by answering exam questions.

Name	Objective	Curriculum and activities
Session 11: Percentages	<ul style="list-style-type: none"> • Work out 10%, 5% and their multiples of any amount without a calculator. • Work out 10%, 5% and their multiples of any amount with a calculator. • Find percentage increase and decreases. 	<p><u>Curriculum</u></p> <p>Using numbers and the number system:</p> <p>3 Multiply and divide whole numbers and decimals by 10, 100, 1000.</p> <p>14 Calculate percentages of quantities, including simple percentage increases and decreases by 5% and multiples thereof.</p> <p>Using common measures, shape and space:</p> <p>18 Calculate simple interest in multiples of 5% on amounts of money.</p> <p>19 Calculate discounts in multiples of 5% on amounts of money</p> <p><u>Purpose of activities</u></p> <ul style="list-style-type: none"> • Review learners understanding of what a percentage is. • Address common misconceptions in finding percentages of an amount. • Find non-calculator methods to find percentages of amounts. • Check current understanding of finding percentages of amounts and to focus learners on different methods and their efficiency. • Highlight different methods. • Model creating ratio tables for percentage questions and to practise using them to support answering questions. • Practise interpreting similar looking questions to identify the key information and to use efficient methods to solve. • Apply the learning to exam questions. • Review lesson objectives and summarise key learning points.

Name	Objective	Curriculum and activities
Session 12: Probability	<ul style="list-style-type: none"> Describe the likelihood of events using a probability scale. Calculate simple probability as a fraction, decimal or percentage, including from a table. Identify and correct common misconceptions around probability. 	<p>Curriculum</p> <p>Handling information and data:</p> <p>30 Understand probability on a scale from 0 (impossible) to 1 (certain) and use probabilities to compare the likelihood of events.</p> <p>31 Use equally likely outcomes to find the probabilities of simple events and express them as fractions</p> <p>Purpose of activities</p> <ul style="list-style-type: none"> Introduce the key language and concept of probability. Review knowledge of fraction, decimal and percentage equivalences. Find probabilities of equally likely events. Locate the probability of simple events and mutually exclusive events on a probability scale. Calculate probability from data in a table. Attempt a selection of exam questions to practise exam technique.
Session 13: Properties of shapes and perimeter	<ul style="list-style-type: none"> Draw 2-D shapes and demonstrate an understanding of line symmetry. Find the missing side lengths in rectilinear and compound rectilinear shapes. Explore the concept of perimeter and calculate the perimeter of compound rectilinear shapes. Understand the concept of perimeter and use it in a range of problem-solving situations. 	<p>Curriculum</p> <p>Using common measures, shape, and space:</p> <p>20 Convert between units of length, weight, capacity, money and time in the same system.</p> <p>Recognise and make use of simple scales on maps and drawings.</p> <p>22 Calculate the area and perimeter of simple shapes including those that are made up of a combination of rectangles</p> <p>Purpose of activities</p> <ul style="list-style-type: none"> Identify characteristics and properties of key shapes. Explore the concept of symmetry. Address common misconceptions around perimeter. Define the concepts perimeter, area, and volume and to recognise which measure should be used. Introduce learners to problems involving compound rectilinear shapes with missing side lengths. Solve multi-step problem-solving questions involving perimeter of compound shapes and money. Check and consolidate their understanding by answering exam questions. Review lesson objectives and recap key learning points.

Name	Objective	Curriculum and activities
Session 14: Area	<ul style="list-style-type: none"> Explore the area of squares, rectangles and compound rectilinear shapes. Understand the concept of area and use them in a range of problem-solving situations. 	<p><u>Curriculum</u> Using common measures, shape, and space: 20 Convert between units of length, weight, capacity, money and time in the same system. 22 Calculate the area and perimeter of simple shapes including those that are made up of a combination of rectangles</p> <p><u>Purpose of activities</u></p> <ul style="list-style-type: none"> Introduce the concept of area. Explore area and highlight the link between area, side lengths and factors. Allow learners to explore area of compound rectilinear shapes. Solve multi-step problem-solving questions involving area of compound shapes, capacity and money. Discuss the process involved in answering multi-step problem-solving exam questions on area. Check and consolidate their understanding by answering exam questions. Summarise learning and review the concept of areas of various shapes.
Session 15: Volume and capacity	<ul style="list-style-type: none"> Define capacity and volume. Convert between measures of capacity and volume. Calculate the volume of cubes and cuboids. Use knowledge of volume and capacity to solve problem-solving questions. 	<p><u>Curriculum</u> Common measures, shape, and space: 20 Convert between units of length, weight, capacity, money and time in the same system. 23 Calculate the volumes of cubes and cuboids.</p> <p><u>Purpose of activities</u></p> <ul style="list-style-type: none"> Highlight the properties of cubes and cuboids and to introduce the context. Work out the volume of a container and discuss the difference between volume and capacity. Identify the cross section of a shape and understand how it is used in the formula for calculating volume Explore sizes of different cuboid containers with constant capacity. Discuss capacity of containers not completely full. Check and consolidate understanding. Recap key learning points and revisit learning objectives.

Appendix 3: Costs

Intervention costs

Providers that signed up to a Multiply Education Research Trial were eligible to receive funding based on the number of learners and the number of additional guided learning hours per trial. The provider funding allocation was calculated as set out below.

Funding area	Funding level and notes
Admin premium	10% premium of the planned learner numbers (based on intervention costs) to support learner recruitment, submission of trial data, trial training and travel expenses.
Intervention costs	£7.20 per additional guided learning hour. Based on number of additional guided learning hours per trial and recruited number of learners to support delivery of the intervention. Plus, area cost and disadvantage cost uplifts will be added, as applicable.
Additional benefits	Each provider received a £1000 one-off payment. Tutors received free, high-quality training on how to implement the lessons.

Control group

Funding area	Funding level and notes
Admin premium	10% premium of the planned learner numbers (based on intervention costs) to support learner recruitment and submission of trial data.
Additional benefit	Each provider received a £1000 one-off payment.

The final grant funding spend for the three-armed trial – involving both Adapted Mastery Approach and Contextualised Approach – was £206,320, according to the funding allocation above. The training cost was £106,870 for the Adapted Mastery Approach and £99,650 for Contextualised Curriculum. This resulted in a total delivery spend for these interventions of £412,840.¹⁷

The product developer for the Adapted Mastery Approach and Contextualised Approach was the Education and Training Foundation (ETF). ETF's deliverables included developing all training and course materials in line with FSQ Level 1 maths, delivering

¹⁷ Figures rounded to the nearest £10

training to tutors, being available for feedback and questions, and delivering the ongoing lesson study sessions to tutors.

Appendix 4: IPE research sub-questions and tools

1. Has the intervention been delivered with fidelity (that is, in line with the intervention guidance)?
2. Is the pre-delivery CPD training delivered as intended, and is it consistently delivered?
 - a. How much of the pre-delivery CPD training do tutors attend?
 - b. To what extent do tutors access the ongoing support throughout delivery of the intervention?
 - c. To what extent do tutors implement the content, principles, approaches and activities in their teaching? How did they do this?
 - d. To what extent do tutors adapt the approaches and activities? What adaptations were made and what were the reasons for this? How far do these adaptations deviate from the prescribed curriculum?
 - e. How feasible is it for tutors to implement the intervention as intended? Where this has not happened, what are the reasons for this?
3. To what extent, if at all, do BAU Functional Skills Qualification (FSQ) Level 1 courses typically include any content relating to the Adapted Mastery Approach (AMA)?
4. Is there any evidence of contamination of BAU FSQ Level 1 courses, or contamination of one of the treatment arms with the other treatment arm? If so, what are the causes of this?
 - a. How, if at all, similar are usual teaching practices for Level 1 FSQ Maths to AMA to FSQ Level 1 teaching practices? How does this vary between providers/tutors?
 - b. To what extent, if at all, do Level 1 FSQ tutors use AMA content and principles in FSQ Level 1 teaching?
 - c. To what extent do we observe contamination of each of the treatment arms with the intervention of the other treatment arm (especially when providers have expressed interest in either intervention)?
5. How did tutors experience delivering the intervention?
 - a. To what extent do tutors understand the rationale behind the interventions? To what extent did they agree with this rationale?
 - b. How do tutors feel about being asked to deliver AMA? What was the buy-in among tutors?
6. To what extent do tutors integrate AMA principles, approaches and activities into their lessons? To what extent did they make changes to their lesson plans? Did this change over time?

- a. How do tutors feel about the integration of AMA into their courses? How well does the format work?
 - b. How easy or difficult is it to deliver and integrate the interventions into the existing curriculum and course hours?
 - c. Are any of the principles, approaches and activities easier or harder to deliver than others?
 - d. To what extent is delivery substantially different to tutors' usual practices?
 - e. How did tutors feel about taking part in a trial? In particular, how did tutors delivering the BAU approach feel about this?
 - f. What, if anything, do tutors feel they gained from the experience of delivering the intervention? How did these outcomes emerge and evolve?
 - g. To what extent do perceived outcomes vary by the characteristics of the tutor, for instance years of experience, or by institutional context and the degree to which the intervention receives strategic support?
 - h. How, if at all, do tutors feel the intervention impact learners?
 - i. How, if at all, could the intervention be further developed or improved to enable wider rollout?
7. What are tutors' experiences of the training and support provided to deliver the intervention?
- a. To what extent do tutors feel the training, tools, resources and ongoing expert support enable them to deliver the interventions effectively? What could be improved?
 - b. How useful were each of those support measures? What worked well and less well? What could be improved?
 - c. What, if anything, do tutors gain from completing the training? (e.g., increased buy-in, understanding of benefits, knowledge and confidence in applying the approach to their teaching)
 - d. Do tutors believe the approach can or should be used in other maths courses in their institutions? Why/why not?
8. What are learners' experiences of the intervention?
9. What motivates learners to take or re-take the FSQ Level 1 teaching course? What are they hoping to gain?
- a. How engaged are learners with the course? How engaged are they in lessons and activities using the AMA?
 - b. What do learners think worked well and less well about the course? What could be improved?
 - c. Further questions for AMA:

- To what extent are learners aware of and understand AMA principles and approaches, and how are they being used?
- How easy or difficult was it to put AMA principles into practice for learners?

10. What outcomes does the intervention have for learners?

- a. What, if anything, do learners feel they gained from the AMA principles, approaches and activities?
- b. How does the intervention, and specifically the adoption of AMA principles, make a difference to these perceived outcomes, including improving learners' ability to make connections between topics?
- c. Does the intervention create interest in taking up a further maths course?
- d. Does the intervention lead to unplanned/unanticipated outcomes?
- e. Do learners gain a greater willingness to try different types of questions and problem solving tasks as a result of the intervention?
- f. What elements were most useful, and were perceived to make the most difference?
- g. To what extent do learner outcomes vary by individual characteristics?

11. What enablers and barriers are there to learner engagement and participation in the intervention?

12. What lessons have been learned from future delivery of the intervention or wider rollout?

- a. Is it plausible the intervention can be rolled out with fidelity on a larger scale?
- b. What, if anything, should be updated or changed for future delivery or wider rollout?

13. What lessons can be learned from the pilot about delivering RCTs in adult education settings?

- a. What can be learned from the process of developing the intervention to be tested through the pilot? What measures were taken to ensure it was suitable for trial? What are the lessons?
- b. What types of promotional activities and messaging worked well / less well to engage and recruit adult education providers to take part in the pilot trial? How / did this differ by different types of education providers?
- c. What motivated adult learning organisations to get involved in the pilot? What concerns did they express for involvement and how were these managed?
- d. What worked in converting adult learning organisations who expressed an interest in participating in the pilot trial to full sign up? What were the enablers / barriers to organisations signing up?

- e. How well did adult education providers understand the randomisation process and were willing to accept the results of this?
 - f. How far were providers compliant with the data collection and sharing requirements associated with the trial? What were the enablers / barriers to effective data collection and sharing? What are the lessons?
 - g. To what extent were participants able to follow the study protocol? What were the enablers / barriers to adherence to this? What are the lessons?
14. Were the outcome measures used valid, reliable, acceptable and complete? How / in what ways could they be improved?
15. What were the reasons for any attrition from the pilot trial? How did attrition vary between treatment and control groups and different types of learning organisations / learners?

Learner interview topic guide

Introduction

Thank you for agreeing to participate in an interview about your experiences of your recent functional skills Level 1 maths course.

I work for King's College London, a public research university and higher education institution. Along with Ipsos and the Institute for Employment Studies (IES), the Department for Education has appointed our research team to evaluate the Adapted Mastery Approach for teaching Level 1 maths functional skills qualification.

Today's discussion will cover your motivations for studying Level 1 maths, your experiences of your recent Level 1 maths course, any outcomes you think came from it, and the barriers and enablers to participating in your course. The interview will last 30 - 45 minutes.

Participation is voluntary, and you can stop at any time. If you do not want to answer a question, let me know and we can move on. Before we begin, I want to run through some general information with you:

- Everything discussed in the interview will only be used for this research. The information you provide will only be viewed by the research team.
- We will write a report based on our findings for the Department for Education. The information you share today will be anonymised in the report. Please feel free to answer the questions as openly and honestly as possible
- With your permission, I would like to record the interview, which would then be transcribed – I'll take notes as we talk, but it helps to have a backup. The recording and transcription will be stored securely and will only be accessible to the KCL team. It will be deleted six months after the end of the project.
- Do you have any questions?
- Are you okay with being voice recorded? [Turn off cameras, turn on recording – announce when recording has started]
- Please confirm that you understand what the research is about, how your data will be used and stored, and that you are happy to take part in the interview.

Background

1. What was your experience of learning maths prior to taking the recent Level 1 functional skills maths course?
 - *Probe for any previous positive/negative experiences of maths*
2. What motivated you to take the Level 1 FSQ maths course?
3. What were you hoping to gain from taking part in the course?

Experience of course

4. What was your overall impression of the course?

[probe on delivery mode, quality of teaching, volume of content, how did it compare to original expectations]

5. Which parts of the course, if any, did you like the most? Why?

6. What parts of the course, if any, did you dislike the most? Why?

7. How, if at all, did this course differ to other maths courses you've taken before?

8. To what extent do your tutor help you understand the course content? How?

9. What kind of approaches did your tutor use to help you understand the course content?

10. As relevant from response, probe key elements of both approaches to check for spillovers.

Maths Mastery approach:

- *Giving you enough time to fully understand a topic before moving on*
- *Breaking ideas in small steps to show how they fit together and build into the larger concept and across the curriculum*
- *Showing the same idea in different ways to help you understand underlying maths principles*
- *Valuing your existing maths knowledge and helping you build on it*
- *Helping you see how different maths topics connect*
- *Spending time understanding patterns and relationships between maths concepts*
- *Encouraging you to use clear mathematical language to explain your thinking*
- *Developing a culture in which everyone believes they can succeed*

11. Which, if any, of the approaches helped you to understand the maths content? Why?

12. Did any of the approaches not work as well for you? Which ones and why?

13. To what extent did your tutor use real life examples?

14. How familiar were the real-life examples that your tutor used in the sessions?

15. In what ways, if any, did the real-world examples help how you engaged with those lessons?

16. Did you attend all your course sessions?

- *Probe: for barriers/facilitators to attending sessions*
- *Probe: were there certain sessions that were more/less engaging? Check for Mastery approaches*

17. Did you complete all the coursework? Probe for barriers – related to broader factors or course itself?

- *Probe: were there certain activities that were more/less engaging? Check for Mastery approaches*

Outcomes

18. What, if anything, did you gain from your Level 1 maths course?

- *Probe for:*
 - *Improved maths outcomes*
 - *Improved understanding of maths language*
 - *Understanding of maths structures*
 - *Better understanding of connections between maths topics*
 - *Confidence*
 - *Increased motivation*
 - *Willingness to attempt maths problems*
 - *Deeper understanding and retention of knowledge*
 - *Improved skills for employment and life tasks*

19. For each outcome referenced, probe on which elements of the course contributed / what their view is based on / whether they have any examples.

20. What, if anything, did you learn in the course that you didn't expect?

21. What, if any, negative impacts or challenges did you experience during the course?

22. How, if at all, has participating in this course affected your interest in continuing with further maths courses?

Enablers and barriers to participation

23. What challenges, if any, did you experience in accessing or completing your course?

24. In what ways do you think the course could be improved?

Closing

Anything else to add before I turn off the recording?

[Turn off recording, turn on camera]

Confirm details for receiving the incentive / gift voucher.

Ask for participants email address for the voucher -- the one we've been using?

Thank you and close

Tutor interview topic guide (treatment – wave 1)

Introduction

Thank you for agreeing to take part in an interview about Adapted Mastery Approach for the level 1 functional skills qualification.

I work for King's College London, a public research university and higher education institution. Along with Ipsos and the Institute for Employment Studies, our research team has been commissioned by the Department of Education to evaluate the Maths Mastery approach for teaching level 1 maths functional skills qualification as part of a pilot randomised controlled trial (RCT).

The evaluation aims to assess the feasibility of implementing the Maths Mastery approach within the functional skills level 1 maths course and the effectiveness of the teaching methods and strategies. It aims to understand what methods are effective in improving adult numeracy skills.

Today's discussion will cover your experiences of Maths Mastery approach training, support and implementation, your early experiences of lesson delivery and your experiences of being part of this trial. The interview will last max 30 minutes.

Participation is voluntary, and you can stop at any time. If you do not want to answer a question, let me know and we can move on. Before we begin, I want to run through some general information with you:

- Everything discussed in the interview will be treated in confidence and only used for the purposes of this research. The information you provide will only be viewed by the research team.
- We will write a report based on our findings for the Department for Education. The information you share today will be anonymised in the report. Please feel free to answer the questions as openly and honestly as possible.
- With your permission, I would like to record the interview, which would then be transcribed – I'll take notes as we talk, but it helps to have a backup. The recording and transcription will be stored securely and will only be accessible to the IES/KCL team. It will be deleted 6 months after the end of the project: September 2026.
- Do you have any questions?
- Please confirm that you understand the research, how your data will be used and stored, and that you are happy to participate in the interview.

Section A: Background

1. Can you very briefly describe the type of organisation you work in?

2. How long have you taught the level 1 functional skills maths qualification (FSQ)? How many years have you been teaching overall?
3. What is the duration of the FSQ course you teach? Probe 15, 17, 30 weeks or other
4. Before the training, had you heard about the Maths Mastery principles and approach, and what experiences, if any, did you have using the principles and the approach in your teaching?
5. How similar, if at all, are 'normal' teaching practices for functional skills maths compared to Maths Mastery approaches and principles?
6. How different, if at all, are 'normal' teaching practices for functional skills maths compared to Maths Mastery approaches and principles?

Section B: Training and Resources

Researcher note: professional development training includes 10 hours spread over 4 sessions. The training covers the Maths Mastery approach, its rationale, and the scheme of work. There are 15 lessons for Mastery, each lasting 1.5 hours, to be delivered to learners over a 15 or 30-week FSQ course. Ongoing support sessions and one-to-one sessions to assist tutors during delivery are mandatory and there is an expectation tutors will attend 80% of these sessions. The support sessions will be around two hours each, and there will be three hours of individualised reflection and coaching support.

7. We understand you attended [use sampling spreadsheet to determine attendance] amount of training. Is this correct?
8. *[if tutors didn't attend all the training sessions]:* What were the main reasons for not attending all the training? Are there any changes or support you think could have encouraged your attendance?
9. *[If they attended some training sessions or none]:* How was information for missed sessions shared with you, if at all?
10. What did you think of the training sessions you attended? Probe for views on content, format, and interaction etc
11. What was most useful about the training? What was least useful?
12. Is there anything you think could be changed or improved about the training?
13. Did you feel the training gave you a good understanding of the Maths Mastery approach and teaching strategies? *If not*, which aspects were you less certain about?
14. How would you briefly describe the Maths Mastery approach and teaching strategies?

Information about elements for researcher, but don't necessarily probe unless relevant:

- *Emphasising a deeper understanding of mathematical concepts*
- *Change of approach from 'covering the curriculum' to deepening understanding and spending time building from context to abstract reasoning*
- *Teaching that allows learners to develop an understanding of mathematical structure.*
- *Valuing and building on learners' prior learning*
- *Prioritising curriculum coherence and connections*

- *Developing both fluency and understanding of key ideas*
- *Developing a culture in which everyone believes everyone can succeed.*

Section C: Post-training

Researcher note: lesson studies are weekly led by a training. The sessions give tutors the opportunity to work together to reflect on teaching practices, deconstruct lesson content, plan for teaching of upcoming lessons and consider the practicalities and possible challenges of delivering the lessons.

Lesson study sessions

15. Are you attending the ongoing weekly lesson study sessions? *Check briefly whether they attended all the lesson study sessions. If not, which ones did they attend?*
16. What do you think of the sessions you have attended so far? *Probe for content, format, reflective practice, lesson deconstruction, lesson delivery challenges and practicalities*
17. To what extent are you finding these session(s) useful? Why/why not?

Other support

18. Have you had any contact with the training team at ETF for additional support outside the training or lesson study sessions?
19. If no:
 - Do you know who to contact if you need additional support?
20. If yes:
 - What did you discuss? Did you ask for help with anything?
21. How useful was the support received?

Overall support

22. Overall, do you feel the training, lesson plans, scheme of work and lesson study sessions have prepared you for delivering the Maths Mastery sessions? *If not, what would have helped you feel more prepared?*

We will contact tutors again once the courses have been delivered. Would you be willing to take part in an interview then?

Anything else to add?

Thank and close

Tutor interview topic guide (treatment – wave 2)

Introduction

Thank you for agreeing to participate in an interview about the Adapted Mastery Approach for the functional skills qualification.

I work for King's College London, a public research university and higher education institution. Along with Ipsos and the Institute for Employment Studies, the Department for Education has appointed our research team to evaluate the Adapted Mastery Approach for teaching level 1 maths functional skills qualification as part of a pilot randomised controlled trial (RCT).

The evaluation aims to assess the feasibility of implementing the Adapted Mastery Approach within the functional skills level 1 math course and the effectiveness of the teaching methods and strategies. It aims to understand what methods are effective in improving adult numeracy skills.

Today's discussion will cover your experiences of the Mastery approach lesson delivery, learner experiences of the programme, outcomes, enablers and barriers to engagement and participation, and your experiences of being part of trial and evaluation. The interview will last 45 minutes.

Participation is voluntary, and you can stop at any time. If you do not want to answer a question, let me know and we can move on. Before we begin, I want to run through some general information with you:

- Everything discussed in the interview will only be used for this research. The information you provide will only be viewed by the research team.
- We will write a report based on our findings for the Department for Education. The information you share today will be anonymised in the report. Please feel free to answer the questions as openly and honestly as possible.
- With your permission, I would like to record the interview, which would then be transcribed – I'll take notes as we talk, but it helps to have a backup. The recording and transcription will be stored securely and will only be accessible to the IES/KCL team. It will be deleted six months after the end of the project. Do you have any questions?
- Please confirm that you understand what the research is about, how your data will be used and stored, and that you are happy to take part in the interview.

Section A: Background (if not interviewed in Wave 1)

1. How long have you taught the level 1 functional skills qualification (FSQ)?
2. *What is the duration of the FSQ course that you teach? Probe 15, 30 weeks or other

3. *Before the training, what had you heard about the...principles and approach, and what experiences, if any, did you have using the principles and the approach in your teaching
 - **Maths Mastery:** *Probe: focus on deep understanding of mathematical principles, prioritising curriculum coherence and connections, developing fluency and understanding of key ideas, building a culture in which everyone believes everyone can succeed*
4. How similar or different are your usual teaching practices for functional skills maths compared to:
 - **Maths Mastery** approaches and principles?

Section B: Delivery

Lesson study sessions

5. *Did you (continue attending) attend the lesson study sessions? Check briefly whether they attended all the lesson study sessions. If not, which ones did they attend? What were the barriers to attendance, and what, if anything, could have enabled them to attend?
6. *What did you think of these sessions? Probe for content, cohort format, reflective practice, lesson deconstruction, lesson delivery challenges and practicalities
7. *Is there anything you think could be changed or improved about these sessions?

Delivery

[Note to researcher – Maths Mastery: Lessons take place within existing curriculum hours. The 22.5 hours of lessons can be delivered as part of weekly 1.5 hour sessions over 15 weeks, or divided over 30 weeks]

8. *How have you found delivering the Maths Mastery approach so far?
9. *How and in what ways have you fitted the Contextualised/Mastery principles and teaching strategies into the FSQ level 1 course?
10. How easy/difficult has this been? Examples?
11. *How / in what ways have you fitted Mastery lessons into the FSQ level 1 course scheme of work/curriculum?
12. How easy/difficult has this been? Examples?
13. *Have you delivered all 15 contextualised lessons? *If not, find out how many and reasons why not? Probe: does it fit into the FSQ level 1 course hours?*
14. *Are any teaching strategies and activities easier or more difficult to deliver than others? Examples?
 - *Probe for these elements of the Maths Mastery approach:*

- *Emphasising a deeper understanding of mathematical concepts. Change of approach from 'covering the curriculum' to deepening understanding and spending time building from context to abstract reasoning*
- *Teaching that allows learners to develop an understanding of mathematical structure.*
- *Valuing and building on learners' prior learning*
- *Prioritising curriculum coherence and connections*
- *Developing both fluency and understanding of key ideas*
- *Developing a culture in which everyone believes everyone can succeed.*

15. What has worked well in terms of delivering the lessons?

16. What has worked less well, what have been the main challenges and barriers?

17. *Have you adapted the strategies or lessons to meet particular learners' general class needs (e.g., SEND, ESOL, learners with less confidence, in person vs online, learners with lower starting points)?

- If yes: What were the reasons?

18. What adaptations did you make?

19. How well do you think these have worked? Why do you think that?

20. *To what extent have senior management at your institution supported you in adopting Maths Mastery? *Do you feel this was sufficient? If not, what else could your institution have done to support you?*

Section C: Learner experiences

21. How familiar do you think Maths Mastery was to the learners? Were they aware that you were teaching differently?

22. To what extent did learners demonstrate an understanding of Mastery approaches and how they are being used?

23. How did this affect engagement and learning, if at all?

24. To what extent did you observe learners exploring mathematical structures and making connections between topics?

25. Do you believe this approach gave them a deeper understanding of maths

26. *How engaged are learners with the Mastery approaches, strategies, and lessons?

Section D: Outcomes

27. *What do you think learners have gained from Mastery activities and lessons

- *Probe for:*
 - *Improved maths outcomes*
 - *Understanding of maths structures*
 - *Confidence*
 - *Increased motivation*
 - *Willingness to attempt maths problems*
 - *Deeper understanding and retention of knowledge*
 - *Improved skills for employment and life tasks*

28. *For each outcome referenced, probe on why they think that / what their view is based on / whether they have any evidence / examples.*

29. *Do you think there are any groups of learners to which the contextual/Mastery lessons and strategies are.....? (*Probe for ESOL, SEND etc*).

- more suited to?
- less suited to?

30. Why do you think that / what is this view based on?

31. *Did the Mastery strategies and approaches lead to any unexpected outcomes for learners?

32. Do you think that the Mastery teaching practices and strategies have impacted how you teach functional skills at level 1? If so, how and in what ways?

33. Will you continue to use them when teaching level 1 FSQ maths? Why / why not?

34. Did you discuss the Mastery approach and teaching strategies with other colleagues at your institution?

35. What were their opinions of the approach and strategies?

36. Have they used any approaches or teaching strategies in their maths teaching? Why / why not? Probe for the type of strategies they have used.

37. Have the learners any unintended or negative effects from participating in Maths Mastery?

Section F: Tutors 'experiences so far of being part of a trial

38. *How and when did you receive information about this trial? Tell us more about that communication.

39. *What are your feelings about being part of the trial? What motivated you to participate, did you feel you had a choice in the decision?

40. *How are you finding the expectations around data collection and evaluation, such as administering learner surveys? What could have made this process easier for you?

41. How clear have the communication been regarding supporting the trial, mainly the information in the trial-readiness packs? What aspects worked well, and what could have been improved?

Section G: Lessons for Future Delivery

42. Would you recommend other adult learning providers to use the Mastery approach for teaching FSQ? (Why/why not?)

43. What could be improved for future delivery if the Mastery approach was offered to other institutions?

44. What are your thoughts on participating in future research or trials like this? Why would you want to be involved, or why not?
45. How do you think we could encourage adult education tutors and the broader sector to participate in more trials or RCTs like this one? What approaches do you believe would work well?

Anything else to add?

Thank and close

Tutor interview topic guide (control)

Introduction

Thank you for agreeing to participate in an interview about the functional skills qualification.

I work for King's College London. Along with Ipsos and the Institute for Employment Studies, the Department for Education (DfE) has appointed our research team to understand what improves adult numeracy.

The DfE is investing in a high-profile trial programme through the Multiply programme to test the effectiveness of different teaching methods and approaches through a pilot randomised controlled trial. Adult education providers from across England, including adult education colleges, independent learning providers, community learning providers, and local and combined authorities, have been recruited to participate in the trials.

Today's discussion will cover your experiences of delivering the functional skills level 1 maths qualification, learners' experiences of the course and your experiences of the trial.

Your provider was in the control group, so you did not receive support to deliver the Adapted Mastery Approach or Contextualised Curriculum programme, but it is still very important to hear about your approaches during the last year, to enable us to interpret the findings of the trial.

The interview will last 30 minutes.

Participation is voluntary, and you can stop at any time. If you do not want to answer a question, let me know and we can move on. Before we begin, I want to run through some general information with you:

- Everything discussed in the interview will only be used for this research. The information you provide will only be viewed by the research team.
- We will write a report based on our findings for the Department for Education. The information you share today will be anonymised in the report. Please feel free to answer the questions as openly and honestly as possible
- With your permission, I would like to record the interview, which would then be transcribed – I'll take notes as we talk, but it helps to have a backup. The recording and transcription will be stored securely and will only be accessible to the IES/KCL team. It will be deleted six months after the end of the project. Do you have any questions?

Turn on recording

- Please confirm that you understand what the research is about, how your data will be used and stored, and that you are happy to take part in the interview.

Section A: Background

1. Can you describe the size and structure of your organisation?
2. *How would you characterise the ethos of your organisation, particularly regarding your commitment to research and evaluation?
3. What is the catchment area for your institution?
4. Can you describe the typical backgrounds of your adult learners?
5. How many years have you been teaching?
6. *How long have you taught the level 1 functional skills maths qualification (FSQ)?
7. *What is the duration of the FSQ course you teach? *Probe 15, 17, 30 weeks or other*
8. *Do you teach any other types of maths courses? *Probe for age of learners*

Section B: Delivery

9. Have you heard about the **Contextualised Maths** principles and approach? If so, what experiences, if any, have you had using the principles and the approach in your teaching?
 - *Probe using real-life contexts, learner-driven models, and tutor facilitating learner progress – i.e., key aspects of the Contextualised maths.*
10. Had you heard about the **Maths Mastery** principles and approach? If so, what experiences, have you had, if any, using the principles and the approach in your teaching?
 - *Probe: focus on deep understanding of mathematical principles, prioritising curriculum coherence and connections, developing fluency and understanding of key ideas, building a culture in which everyone believes everyone can succeed.*
11. Tell me a bit about how you deliver functional skills.
12. To what extent do you use any real-world contexts to teach FSQ maths? If so, how do you use them in your teaching?
13. To what extent do you emphasise building deeper understanding of mathematical concepts and structures, and how they are interconnected when you teach FSQ maths?
14. What is important to keep learners engaged with maths?
15. What do adult learners struggle with within the functional skills curriculum?

16. What motivates adult learners to take or re-take the FSQ level 1 teaching course?
What do you think learners hope to gain?
17. What worked well / less well about delivering the FSQ course? What could be improved?

Section C: Outcomes

18. How engaged are learners with the FSQ lessons?

- *Probe for:*
 - *Attendance*
 - *Willingness to participate in class*
 - *Attitudes towards the FSQ course*
 - *Willingness to attempt maths problems*

19. How / do you think the FSQ Maths Level 1 activities and lessons have impacted learners? What is this based on? Do you have any evidence / examples you can share?

- *Probe for:*
 - *Understanding of maths concepts*
 - *Confidence in maths*
 - *Motivation for maths*
 - *Problem-solving skills*
 - *Deeper understanding and retention of knowledge*
 - *Improved skills for employment and life tasks*

20. *Do you feel there are any groups of adult learners which the FSQ maths level 1 activities and lessons are...? (*Probe for ESOL, SEND etc*).

- more suited to?
- less suited to?

Section D: Enablers and barriers to participation

21. Have there been any challenges or barriers to delivering the FSQ maths curriculum and lessons?

- *Probe for:*
 - *Course content*
 - *Staffing*
 - *Learner attendance*
 - *Learner barriers – e.g. SEND, ESOL (speech and language barriers)*
 - *Learner attitudes or confidence*

Section E: Tutors' experiences so far of being part of a trial

22. *How and when did you receive information about this trial, tell us more about that communication?
23. *What are your feelings about being part of the trial? What motivated you to participate, and did you feel you had a choice in the decision?
24. *How are you finding the expectations around data collection and evaluation, such as administering learner surveys? What could have made this process easier for you?
25. How clear have your communications been regarding supporting the trial, particularly the information in the trial-readiness packs? What aspects worked well, and what could have been improved?
26. Were there any challenges or barriers in being part of this trial?
27. How could these barriers be overcome in the future?
28. *What are your thoughts on participating in future research or trials like this? Why would you want to be involved, or why not?
29. *How do you think we could encourage adult education tutors and the broader sector to take part in more trials or RCTs like this one? What approaches do you believe would work well?

Anything else to add?

Thank and close

Stakeholder interview topic guide

Introduction

Thank you for agreeing to participate in an interview about the Multiply Research Trials. The Multiply programme aims to test the effectiveness of different teaching methods and approaches through a pilot randomised controlled trial (RCT).

I work for King's College London, a public research university and higher education institution. Along with Ipsos and the Institute for Employment Studies (IES), the Department for Education (DfE) has appointed our research team to understand what improves adult numeracy.

Our discussion today will cover your motivations for participating in the Multiply trials, your experiences of being part of this, including what has gone well and less well and why, and recommendations for future trials in the sector. The interview will last around 45 minutes.

Participation is voluntary, and you can stop at any time. Let me know if you do not want to answer a question, and we can move on. Before we begin, I want to run through some general information with you:

- Everything discussed in the interview will only be used for this research. The information you provide will only be viewed by the research team.
- We will write a report based on our findings for the Department for Education. The information you share today will be anonymised in the report. Please feel free to answer the questions as openly and honestly as possible
- With your permission, I would like to record the interview, which would then be transcribed – I'll take notes as we talk, but it helps to have a backup. The recording and transcription will be stored securely and only accessible to the KCL team. It will be deleted six months after the end of the project. Do you have any questions?
- Are you okay with being voice recorded?

Turn off cameras, turn on recording.

- Please confirm that you understand the research, how your data will be used and stored, and that you are happy to participate in the interview.

If provider mentions being a part of the Embedded trail: Ask how it impacts delivery of Maths Mastery that they are also part of the Embedded trial.

Section A: Background

1. Tell me a bit about your role, and how long have you been this role?

2. How long have you worked in education?
3. I understand you were involved in the decision for your organisation to participate in the FSQ trial. **If not**, could you tell me who the decision maker was? Can you pass on their contact details so we can speak to them about the trial?

[if no – explore their involvement in anything relevant to the rest of the topic guide. If no relevance, close interview and contact relevant contact]

Section B: Trial Communications

4. How and when did you hear about the trial?
5. Why did your institution decide to participate?
6. Did you consult with the teaching staff when you were making the decision to be involved? Why/Why not?
7. After signing up for the trial, how and when did you receive further information about the trial? Tell us more about that communication and how it worked from your perspective.
8. Did you receive a trial-readiness pack? If yes, what were your overall thoughts on the information included?
 - *Prompts:*
 - *Did the pack provide all the necessary details about the training and evaluation requirements?*
 - *What information, if anything, do you wish had been included?*
 - *Did any information seem unnecessary or irrelevant?*
9. Overall, how clear or unclear have the communications been regarding participating in the Mastery trial, including the training and evaluation requirements?
 - *Prompts:*
 - *What aspects worked well?*
 - *What could have been improved?*
10. Did you share the trial readiness pack for tutors with your staff?
11. Was this helpful in addressing their information needs? Why/why not?
12. Did they understand what training would be involved and their role in the evaluation? Why do you say that?
13. Did you run additional information sessions for tutors? Any other support?
14. In what ways has your institution disseminated information to learners about the trial? How has your institution ensured learners have sufficient information to decide to be involved?
 - *Prompts:*

- *Did you /your staff talk them through the learner information during enrolment?*
- *Did you / your staff give them copies of the learner information?*

Section C: Trial Delivery

15. Who in your organisation is responsible for uploading the information about tutors and learners to the secure portal supplied by Ipsos? How did they find/are they finding uploading the information to the portal?
16. Have you contacted any of the evaluation teams for support with the evaluation activities (Ipsos, IES, KCL)?
- If yes, what did you discuss? What was helpful / less helpful about this discussion?
 - If no, were you aware that you could contact the evaluation teams for support?
17. Have you or your staff participated in Etio's monthly keeping-in-touch calls?
- If yes, how did you find these trial support calls?
18. What was discussed?
19. What aspects of the call were helpful, and what was less helpful?
- If not, were you aware that Etio was hosting monthly keeping-in-touch calls?
20. Did you receive a compliance visit from Etio?
- If yes, how did that go? Probe: how made aware of visit, how did they prepare, access to Etio materials
21. What was helpful / less helpful about the visit?

Section D: Training and Resources

22. How were staff released for training, and did this align with their contracts and timetables, or did it require additional cover? (probe: released from contract hours? Resources/adaptations to enable this?)
23. Did you feel the time for training and lesson study sessions were realistic and feasible?
24. What feedback did you receive from staff regarding training expectations for the study? (i.e., attendance at 3 training sessions (10 hours) and 80% of 12 weekly study sessions.) Why do they say that?
25. Did you attend any of the training sessions? What did you think of the training?
- **If 'yes' – they also attended the training**

26. Did the training give you a good understanding of the Mastery maths approaches?
Why do you think that?
27. Did you attend any lesson study sessions?
28. What did you think of these? What feedback, if any, did you receive from staff participating in the lesson study sessions? Where are these helpful?
29. Is there anything that could be changed or improved about these sessions?

Ask ALL

30. How did your institution decide which tutors would deliver Mastery approach following the training?
31. If staff had questions about delivering Mastery approaches, do you know if they had any contact with the training team at the Education and Training Foundation (ETF) for additional support outside the training or lesson study sessions?
32. What did they discuss?
33. Did they know who to contact?
34. Did they feel they could contact ETF with queries?
35. How have staff received the Maths Mastery approach for FSQ level 1? Have they been interested, engaged or motivated by approach? Why do you say that?

Section E: Outcomes

36. In what ways, if at all, do you think the Mastery Approach has influenced staff teaching methods for FSQ Maths Level 1 and their confidence levels? Can you share any specific examples of these impacts?
37. What signs, if any, have you noticed that suggest a Mastery approach is becoming more common within your organisation?
38. Have any unintended or negative effects resulted from participating in the Mastery approach for teaching FSQ level 1 maths?

Section F: Views of the trial

39. What were the resources needed? Probe for staff time and any other costs and resources
- Facilitation trial
 - Programme itself
40. Based on your experiences, has the programme been worthwhile? Can you share any examples?

41. Were there any challenges or barriers in being part of this programme and trial? If so, what were these barriers/challenges? How could these barriers be overcome in the future?
42. Is there anything you would change to improve the programme and trial? If so, what would that be?
43. Would you recommend that the Mastery approach be rolled out to more adult learning providers? Why/Why not?
44. What are your thoughts on participating in future research or trials like this? Why would you want to be involved, or why not?
45. How could we encourage adult education providers/the broader sector to participate in more trials or RCTs like this? What approaches do you believe would work well?

Anything else to add?

[Turn off recording, turn on camera]

Thank and close

Survey questions about maths confidence

This section lists the survey questions about maths confidence, with detailed overview of answer options. These were asked in both the baseline and endline learner surveys.

Overall, how confident do you feel working with numbers in everyday life?

1. Very confident
2. Fairly confident
3. Neither confident nor not confident
4. Not very confident
5. Not at all confident
6. Not relevant to me

And how confident do you feel about doing the following things in everyday life?

- a) Checking your change is right when you have bought something
- b) Working out the best deals when shopping
- c) (if parental responsibility) Helping children with maths homework or talking about maths / numbers with children
- d) Understanding interest rates on bank statements
- e) Keeping track of your bank account balance
- f) (if working before course start) Working with numbers as part of a job

1. Very confident
2. Fairly confident
3. Neither confident nor not confident
4. Not very confident
5. Not at all confident
6. Not relevant to me

Appendix 5: Code

Randomisation code

The following code was used to randomise providers across the 3 arms. Randomisation took into account randomising through all relevant gateways. The steps are shown following the randomisation of the first batch (or set) of providers:

1. For each new batch of providers, combine it with all previously included providers and rerun the randomisation across the full set of units (treatment/control). Because the same random seed is used each time, the assignment of previously randomised providers remains unchanged. This method generates assignments for the new providers while preserving the original allocations and ensuring consistency with the overall randomisation.
2. Keep original assignment for old batch and include newly randomised assignment for new batch.
3. Repeat steps 1 and 2 for each new batch.

This method utilising `rank()` ensures a more stable randomisation for smaller sample sizes compared to naively randomising just using `runiform() $>x$` .

```
clear
import excel "~\file path"
gen cont=0
//replace cont=1 if LeadId==" " //mark providers set for CC (or MM) only

gen region=.
replace region=1 if Region=="London"
replace region=2 if Region=="East Midlands"
replace region=3 if Region=="South East of England"
replace region=4 if Region=="Yorkshire"
replace region=5 if Region=="North East"
replace region=6 if Region=="East Midlands"
replace region=7 if Region=="East of England"
replace region=8 if Region=="North West"
replace region=9 if Region=="South West"

set seed 585506 ///Sets seed for replication.

forvalues i=1/9 {
tab cont if cont==0 & region==`i'
scalar N_`i'=r(N)/3
disp "`i'=N_`i'" and Region: `i'"
}

//Assign random numbers to the observations and rank them from the smallest to the largest
bysort region: gen random_number_Combined=uniform() if cont==0 // [GENERATES A RANDOM
NUMBER BETWEEN 0 AND 1]
bysort region: egen ordering_Combined=rank(random_number) if cont==0 //[ORDERS EACH
OBSERVATION FROM SMALLEST TO LARGEST]
```

```

//Assign observations to control & treatment group based on their ranks
gen group=.
  forvalues i=1/9 {
    replace group=0 if ordering_Combined <= `=N_`i' & cont==0 & region==`i' //[ASSIGNS TREATMENT
STATUS TO FIRST THIRD OF SAMPLE]
      replace group=1 if ordering_Combined > `=N_`i' & ordering_Combined <=(`=2*N_`i') & cont==0 &
region==`i' //[ASSIGNS CC STATUS TO SECOND THIRD OF SAMPLE]
      replace group=2 if ordering_Combined > (`=2*N_`i') & cont==0 & region==`i' //[ASSIGNS MM STATUS
TO LAST THIRD OF SAMPLE]

//If only one observation per region, ordering will automatically assign to group=2.
//To overcome this, if random_number_Combined>1/3 we randomly assign to MM or CC with 50%
prob. (group=1 or 2), else we assign to control (group=0)
if `=N_`i'==(`=1/3') {
  replace group=0 if random_number_Combined<`=1/3' & cont==1 & region==`i'
  replace group=1 if random_number_Combined>`=1/3' & uniform() <= 0.5 & cont==1 & region==`i'
}
}

//For Contextualised Curriculum
bysort region: gen random_number_Context=uniform() if cont==1 //
bysort region: egen ordering_Context=rank(random_number_Context) if cont==1 //

forvalues i=1/9 {
  tab cont if cont==1 & region==`i'
  scalar N_`i'=r(N)/2
}
forvalues i=1/9 {
  replace group=0 if ordering_Context <= `=N_`i' & cont==1 & region==`i'
  replace group=1 if ordering_Context > `=N_`i' & cont==1 & region==`i'

//If only one observation per region, ordering will automatically assign to group=1.
//To overcome this, we assign to control (group=0) if random_number_Context<=0.5
if `=N_`i'==(`=1/2') {
  replace group=0 if random_number_Context <= `=N_`i' & cont==1 & region==`i'
}
}

```

Analytical code (main, subgroup, missing data, CACE)

```

clear all //Set folder path (user only needs to change this line) global datafolder "C:/your/path/here"

//===== //ASSEMBLE DATA //=====

//Treatment import excel "$datafolder/mathsmastery.xlsx", sheet("trial_data") allstring firstrow clear drop fn
// remove family numeracy variables tempfile treat save `treat'

//Control import excel "$datafolder/fsqlevel1.xlsx", sheet("trial_data") allstring firstrow clear drop fn // drop
family numeracy variables replace providername=lower(providername)

```

```

//remove gateway 3 providers drop if strpos( providername, "thecityofliverpoolcollege") | strpos(
providername, "walthamforestlondonboroughcouncil-adultlearning") | strpos( providername,
"westnottinghamshirecollege")

append using `treat'

//Lock order prior to merging 1:m for postcode detection set seed 2321 gen double _sortorder=runiform()
sort uniquelearnerid _sortorder, stable drop if svy_bs_s_trialdum=="embedded maths" //Remove sample of
embedded maths-only participants

// Define sessions attended (to be used later on as part of CACE) forvalues i=1/15 { gen session_ni=1 if
sessioni=="Attended all" | sessioni=="Attended part" replace session_ni=0 if session`i'=="Did not attend" }
egen tot_sess=rowtotal(session_n*)

// Catch aims and align them with relevant outcome and outgrade outcome variables unab aims:
ilr_learnaimrefitle_aim* gen aims=. gen outcome="" gen outgrade="" forvalues i=1/20 { replace
ilr_learnaimrefitle_aimi=lower(ilr_learnaimrefitle_aimi) replace aims=i' if
strpos(ilr_learnaimrefitle_aimi,"functional skills qualification in math") &
!strpos(ilr_learnaimrefitle_aimi,"level 2") & !strpos(ilr_learnaimrefitle_aimi,"3") replace outgrade=
ilr_outgrade_aimi' if strpos(ilr_learnaimrefitle_aimi,"functional skills qualification in math") &
ilr_outgrade_aimi!="" & !strpos(ilr_learnaimrefitle_aimi,"level 2") & !strpos(ilr_learnaimrefitle_aimi,"3")
replace outcome= ilr_outcome_aimi' if strpos(ilr_learnaimrefitle_aimi,"functional skills qualification in
math") & ilr_outcome_aimi!="" & !strpos(ilr_learnaimrefitle_aimi,"level 2") &
!strpos(ilr_learnaimrefitle_aimi,"3") } gen outcome_n=. replace outcome_n=1 if aims!=. & outgrade=="PA"
replace outcome_n=0 if aims!=. & outcome=="3" replace outcome_n=1 if aims!=. & outcome=="1" gen
tagg=1 if group!="" & outcome_n==. //ITT replace outcome_n=0 if group!="" & outcome_n==. //ITT replace
outcome_n=. if aims!=. & outcome=="8" //Impute Code 8: Learning activities are complete but the outcome
is not yet known

gen compstatus="" gen date="" forvalues i=1/20 { replace compstatus=ilr_compstatus_aimi' if aims==i'
replace date=ilr_learnstartdate_aimi' if aims==i' }

//Flag non-missing aims gen aims_nonmiss=1 if aims!=. gen treat=. replace treat=1 if group=="treatment"
replace treat=0 if group=="control"

//===== // Set up variables //===== encode providerid,gen(prov_id)
encode uniquelearnerid,gen(learn_id)

gen white=0 replace svy_bs_qethnicity=lower(svy_bs_qethnicity) replace white=1 if
strpos(svy_bs_qethnicity,"english") replace white=1 if strpos(svy_bs_qethnicity,"white") replace white=0 if
strpos(svy_bs_qethnicity,"white and")

rename ilr_sex sex gen byte female=1 if sex=="F" replace female=0 if sex=="M" lab define yesno 1 "Yes" 0
"No" .a "Unknown", replace lab val female yesno lab var female "Female" drop sex

destring ilr_ethnicity, gen(ethnicity) gen ethnicity_cat=.

//White replace ethnicity_cat=1 if inlist(ethnicity, 31, 32, 33, 34)

//Mixed or multiple ethnic groups replace ethnicity_cat=2 if inlist(ethnicity, 35, 36, 37, 38)

//Asian or Asian British replace ethnicity_cat=3 if inlist(ethnicity, 39, 40, 41, 42, 43)

//Black, Black British, Caribbean or African replace ethnicity_cat=4 if inlist(ethnicity, 44, 45, 46)

```

```

//Other ethnic group replace ethnicity_cat=5 if inlist(ethnicity, 47, 98)

label define ethnicity_grp_lab 1 "White" 2 "Mixed or multiple ethnic groups" 3 "Asian or Asian British" /// 4
"Black, Black British, Caribbean or African" 5 "Other ethnic group"

label values ethnicity_cat ethnicity_grp_lab

gen ethnicity2=. replace ethnicity2=1 if ethnicity_cat==1 replace ethnicity2=0 if ethnicity_cat!=1 &
ethnicity_cat!=.

gen employed=. replace employed=1 if inlist(ilr_l_empstatfdl,"1","2","3","4","5","6","13") replace
employed=1 if inlist(ilr_l_empstatfdl,"14","15","16","19","20","21") replace employed=1 if
inlist(ilr_l_empstatfdl,"22","23","24","25","26") replace employed=0 if inlist(ilr_l_empstatfdl,"7","8","9","10",
"11","12","17","18") replace employed=. if inlist(ilr_l_empstatfdl,"-1","98") label define employed_lbl 0
"Unemployed" 1 "Employed" label values employed_employed_lbl

destring ilr_llddhealthprob, gen(llddhealthprob) #delimiter; lab define llddhealthprob 1 "Learner reports
Learning Difficulty, Disability or Health Problem" 2 "Learner reports no Learning Difficulty, Disability or
Health Problem" 9 "No Information Provided By The Learner"; #delimiter cr lab val llddhealthprob
llddhealthprob lab var llddhealthprob "Health problem"

gen illhealth=1 if llddhealthprob==1 replace illhealth=0 if llddhealthprob==2 replace illhealth=.a if
llddhealthprob==9 lab val illhealth yesno lab var illhealth "Learner reports health problem"

replace svy_end_fsq_qconf=lower(svy_end_fsq_qconf) local confidence_end svy_end_fsq_qnumconf_aq2
svy_end_fsq_qnumconf_bq2 svy_end_fsq_qnumconf_cq2 svy_end_fsq_qnumconf_dq2
svy_end_fsq_qnumconf_eq2 svy_end_fsq_qnumconf_fq2 local confidence_base svy_bs_qnumconf_aq2
svy_bs_qnumconf_bq2 svy_bs_qnumconf_cq2 svy_bs_qnumconf_dq2 svy_bs_qnumconf_eq2
svy_bs_qnumconf_fq2 local vars_conf svy_bs_qconf svy_end_fsq_qconf local all_conf confidence_end'
confidence_base' vars_conf' foreach v of local all_conf { replace v'=lower(v') gen v'_n=. replace v'_n=5 if
strpos(v',"very") replace v'_n=4 if strpos(v',"fairly") replace v'_n=3 if strpos(v',"neither") replace v'_n=2 if
strpos(v',"not very") replace v'_n=1 if strpos(v',"not at all") }

egen confidence_maths_bs=rowmean(svy_bs_qnumconf_aq2_n svy_bs_qnumconf_bq2_n
svy_bs_qnumconf_cq2_n svy_bs_qnumconf_dq2_n svy_bs_qnumconf_eq2_n svy_bs_qnumconf_fq2_n)
egen confidence_maths_end=rowmean(svy_end_fsq_qnumconf_aq2_n svy_end_fsq_qnumconf_bq2_n
svy_end_fsq_qnumconf_cq2_n svy_end_fsq_qnumconf_dq2_n svy_end_fsq_qnumconf_eq2_n
svy_end_fsq_qnumconf_fq2_n)

gen region=""

replace region="East Midlands" if strpos(providername,"chesterfi") replace region="North East" if
strpos(providername,"darling") replace region="East Midlands" if strpos(providername,"derby") replace
region="East Midlands" if strpos(providername,"northampt") replace region="South West" if
strpos(providername,"theroyal") replace region="East of England" if strpos(providername,"thurrock")
replace region="East Midlands" if strpos(providername,"evolveyourfuturelimited") replace
region="Yorkshire and the Humber" if strpos(providername,"futureworks") replace region="Yorkshire and
the Humber" if strpos(providername,"creatingpositiveopportunity") replace region="East of England" if
strpos(providername,"essexacl") replace region="East of England" if strpos(providername,"luton") replace
region="North West" if strpos(providername,"phxtr") replace region="East of England" if
strpos(providername,"southend") replace region="North West" if
strpos(providername,"womenstechnologytraininglimited") replace region="North West" if

```

```

strpos(providername,"blackpoolboroughcouncil") replace region= "London" if
strpos(providername,"hounslow")

replace region= "Yorkshire and the Humber" if strpos(providername,"kingstonuponhullcitycouncil")

replace region= "East of England" if strpos(providername,"peterborou") replace region= "South West" if
strpos(providername,"petroc") replace region= "Yorkshire and the Humber" if
strpos(providername,"northyorkshirecountycouncil") replace region= "London" if
strpos(providername,"harrow") replace region= "London" if strpos(providername,"thewea") replace region=
"Yorkshire and the Humber" if strpos(providername,"northlincolnshirecounciladulthoodeducatio") replace
region= "North East" if strpos(providername,"trriage") replace region= "Yorkshire and the Humber" if
strpos(providername,"cityofdoncaster") | strpos(providername,"doncaster") replace region= "London" if
strpos(providername,"thecityliter") replace region= "East of England" if strpos(providername,"haveri")
replace region= "East of England" if strpos(providername,"centralbedfordshirecouncil") replace region=
"London" if strpos(providername,"globalskills") replace region= "South East" if
strpos(providername,"hittraining") replace region= "London" if strpos(providername,"islington") replace
region= "London" if strpos(providername,"lewisham") replace region= "London" if
strpos(providername,"londonvocationalcollegelimited") replace region= "South East" if
strpos(providername,"miltonkeynescitycouncil") replace region= "North West" if
strpos(providername,"southport") replace region= "East of England" if
strpos(providername,"westsuffolkcollege") replace region= "North West" if
strpos(providername,"women'stechnologytraininglimited")

encode age, gen(age_n) encode region, gen(region_n)

merge m:1 postcode using "$datafolder/depr_postcodes.dta" keep if _merge==3 | _merge==1 drop _merge

gen deprivation=1 if F<3 & F!=. replace deprivation=0 if F>=3 & F!=.

//Remove duplicate entries sort uniquelearnerid aims_nonmiss _sortorder, stable by uniquelearnerid: keep
if _n== 1

//===== //BALANCE TEST //===== foreach v of varlist female ethnicity2 age_n
illhealth deprivation employed { reg treat i.`v' i.region_n, vce(cluster prov_id) } preserve keep if treat==1
reshape long session_n, i(uniquelearnerid) j(time)

//cumulative dosage per learner bysort uniquelearnerid (time): gen cum_attend=sum(session_n)

//collapse to time level collapse (mean) mean_cum=cum_attend (mean) p_attend=session_n, by(time)

//graph 1: mean cumulative dosage twoway line mean_cum time, lcolor("16 79 117") name(g1, replace)
ytitle("Cumulative Attendance (mean)") xtitle("") xlabel(1(1)15) ylabel(1(2)15, angle(horizontal))

//graph 2: mean session attendance twoway line p_attend time, lcolor("16 79 117") name(g2, replace)
ytitle("Attendance (proportion)") xtitle("Sessions (time)") xlabel(1(1)15) ylabel(0.65(0.05)0.85,
format(%3.2f)) yscale(range(0.65 0.85))

graph combine g1 g2, cols(1) imargin(tiny) xcommon

restore

//===== //SUMMARY STATISTICS //===== local contvars outcome_n
age_n local catvars treat region_n female ethnicity_cat illhealth employed deprivation
svy_end_fsq_qconf_n svy_bs_qconf_n

```

```

tempfile summstats tempname S

postfile S' str25 section str32 var str80 level str20 stat double value using summstats', replace

//continuous foreach v of local contvars { quietly summarize v' post S' ("cont_summary") ("v") ("" ("mean")
(r(mean)) post S' ("cont_summary") ("v") ("" ("sd") (r(sd)) post S' ("cont_summary") ("v") ("" ("n") (r(N)) }

//categorical foreach v of local catvars { preserve keep v' contract v', freq(freq) quietly summarize freq
scalar Ncat=r(sum) gen double percent=100*freq/Ncat

///post each level forvalues i=1/_N' { local lvl "" capture confirm string variable v' if _rc==0 { local
lvl="v'[i]'" } else { local lvl=string(v'[i]') } post S' ("cat_freq") ("v") ("lvl") ("freq") (freq[i]) post S' ("cat_freq")
("v") ("lvl") ("percent") (percent[i]) } restore }

//Raw means of outcome_n by treat foreach t in 0 1 { quietly summarize outcome_n if treat==t post S'
("raw_by_treat") ("outcome_n") ("treat=t") ("mean") (r(mean)) post S' ("raw_by_treat") ("outcome_n")
("treat=t") ("sd") (r(sd)) post S' ("raw_by_treat") ("outcome_n") ("treat=t") ("n") (r(N)) } postclose S'
preserve use `summstats', clear order section var level stat value sort section var level stat save
"summary_results.dta", replace restore

//Save original dataset tempfile touse save `touse', replace

//===== //PROVIDERS INVOLVED + CHARACTERISTICS
//===== local prov "prov_id" local id "learnerid" local treat "treat"
local prov_chars "region_n" local base_cont "age_n deprivation" local base_cat "region_n female ethnicity2
illhealth employed" local pretests "svy_bs_qconf_n confidence_maths_bs" local posttests
"svy_end_fsq_qconf_n confidence_maths_end" local attr_reason "attrition_reason" local xlsx
"report_outputs.xlsx" capture erase "`xlsx'"

///Provider roster preserve keep prov' treat' prov_chars' bys prov': keep if _n==1 export excel using "`xlsx'",
sheet("providers_roster") firstrow(variables) sheetreplace restore

///Provider counts (compute first; then build export dataset) preserve keep prov' treat' bys `prov': keep if
_n==1

count local nall=r(N) count if treat'==0 local n0=r(N) count if treat'==1 local n1=r(N) restore

clear set obs 3 gen stat="" gen value=. replace stat="N_providers_all" in 1 replace value=nall' in 1 replace
stat="N_providers_control" in 2 replace value=n0' in 2 replace stat="N_providers_treat" in 3 replace
value=n1' in 3 format value %9.0f export excel using "xlsx", sheet("providers_counts") firstrow(variables)
sheetreplace

///Provider characteristics distributions by arm (one sheet per var) use touse', clear foreach v of local
prov_chars { preserve keep prov' treat' v' bys prov': keep if _n==1 drop if missing(v) | missing(treat')
contract treat' v', freq(n) gen varname="v" format n %9.0f export excel using "xlsx", sheet("provchar_v")
firstrow(variables) sheetreplace restore }

//===== //BASELINE BALANCE CHECKS //=====
use `touse', clear

tempfile bal postfile hb str40 covar str80 term double b se p N using `bal', replace

foreach v of varlist female ethnicity2 illhealth deprivation employed age_n {

```

```

if "v"=="age_n" { quietly regress treat' c.v' i.region_n, vce(cluster prov') } else { quietly regress treat' i.v'
i.region_n, vce(cluster `prov') }

local N=e(N) matrix V=e(V) local cnames: colfullnames e(b)

///keep only coefficients corresponding to this covariate's levels foreach t of local cnames { if strpos("`t'", "v")
& ("t"!="_cons") { local b=_b[t] local se=sqrt(V[colnumb(V,"t"),colnumb(V,"t")]) local df=e(df_r) local
p=2*ttail(df, abs(b'/se')) post hb ("v") ("t") (b') (se') (p') (^N') } } postclose hb

clear use `touse'

preserve use bal', clear format b se p %9.3f format N %9.0f export excel using "xlsx",
sheet("balance_reg_treat") firstrow(variables) sheetreplace restore

///Categorical covariates (level-by-level) tempfile balcat postfile hbg str40 varname str60 level double
prop_c prop_t diff se p N using `balcat', replace

foreach v of local base_cat {

quietly levelsof v' if !missing(v') & !missing(`treat'), local(levels)

foreach L of local levels { tempvar d gen byte d'=(v'==L') if !missing(v') & !missing(`treat')

quietly summarize d' if treat'==0 & !missing(`d') local pc=r(mean)

quietly summarize d' if treat'==1 & !missing(`d') local pt=r(mean)

quietly regress d' treat' if !missing(d') & !missing(treat'), vce(cluster prov') local b=_b[treat'] local
se=_se[treat'] local df=e(df_r) local p=2*ttail(df, abs(b'/se')) local N=e(N)

post hbg ("v") ("L") (pc') (pt') (b') (se') (p') (N') drop d' } } postclose hbg clear use touse' preserve use
balcat', clear format prop_c prop_t diff se p %9.3f format N %9.0f export excel using "xlsx",
sheet("balance_categorical") firstrow(variables) sheetreplace restore

*===== //PRE-TEST DIFFERENCES *===== tempfile pretres
postfile hp str40 pretest double mean_c mean_t diff se p N using `pretres', replace

foreach v of local pretests {

quietly summarize v' if treat'==0 & !missing(`v') local mc=r(mean)

quietly summarize v' if treat'==1 & !missing(`v') local mt=r(mean)

quietly regress v' treat' if !missing(v') & !missing(treat'), vce(cluster prov') local b=_b[treat'] local
se=_se[treat'] local df=e(df_r) local p=2*ttail(df, abs(b'/se')) local N=e(N)

post hp ("v") (mc') (mt') (b') (se') (p') (N') } } postclose hp clear use touse' preserve use pretres', clear format
mean_c mean_t diff se p %9.3f format N %9.0f export excel using "xlsx", sheet("pretest_differences")
firstrow(variables) sheetreplace restore

*===== //PRE-TEST DISTRIBUTIONS AND HISTOGRAMS
*===== tempfile distsum postfile hd str40 pretest double N
mean sd min p25 p50 p75 max using `distsum', replace

foreach v of local pretests {

```

```
quietly summarize v' if !missing(v'), detail post hd ("`v'" (r(N)) (r(mean)) (r(sd)) (r(min)) (r(p25)) (r(p50))
(r(p75)) (r(max))
```

```
local xtitle "Single measure" if strpos("`v'", "confidence") { local xtitle "Composite measure" }
```

```
capture graph drop h_v' histogram v' if !missing(v'), name(h_v', replace) color(navy) lcolor(navy)
graphregion(color(white)) plotregion(color(white)) xtitle("xtitle") graph export "hist_v'.png", replace }
postclose hd clear use touse' preserve use distsum', clear format mean sd min p25 p50 p75 max %9.3f
format N %9.0f export excel using "`xlsx'", sheet("pretest_distributions") firstrow(variables) sheetreplace
restore
```

```
*===== //ATTRITION RATES *=====
```

```
tempfile attrres postfile ha str40 outcome str40 pretest str40 posttest double N_base N_obs attrit_rate ///
double rate_c rate_t using `attrres', replace
```

```
local k: word count pretests' forvalues i=1/k' {
```

```
local y0: word i' of pretests' local y: word i' of posttests'
```

```
capture drop attrit_y' gen byte attrit_y'=missing(y') if !missing(y0') & !missing(`treat')
```

```
quietly count if !missing(y0') & !missing(treat') local Nbase=r(N)
```

```
quietly count if attrit_y'==0 local Nobs=r(N)
```

```
quietly summarize attrit_y' if attrit_y'<. local ar=r(mean)
```

```
quietly summarize attrit_y' if attrit_y'<. & `treat'==0 local ar0=r(mean)
```

```
quietly summarize attrit_y' if attrit_y'<. & `treat'==1 local ar1=r(mean)
```

```
post ha ("y") ("y0") ("y") (Nbase') (Nobs') (ar') (ar0') (ar1')
```

```
if "attr_reason" != "" { capture confirm variable attr_reason' if !rc { preserve keep if attrity'==1 keep treat'
attr_reason' drop if missing(attr_reason') contract treat' attr_reason', freq(n) gen outcome="`y'" format n
%9.0f
```

```
local sh="attr_reasons_i" export excel using "xlsx", sheet("sh") firstrow(variables) sheetreplace restore } }
} postclose ha clear use touse' preserve use attrres', clear format attrit_rate rate_c rate_t %9.3f format
N_base N_obs %9.0f export excel using "xlsx", sheet("attrition_rates") firstrow(variables) sheetreplace
restore
```

```
//===== //Missingness Prediction //=====
```

```
tempfile miss_level_pvals tempname P
```

```
postfile P' str32 outcome str40 term double N double coef double se double z double pval using
miss_level_pvals', replace
```

```
local m confidence_maths_end svy_end_fsq_qconf_n outcome_n foreach y of local m {
```

```
///Create missingness indicator (overwrite safely) capture drop y'_miss gen byte y'_miss=missing(`y')
```

```
///Fit model
```

```

logit `y'_miss i.treat i.female i.ethnicity_cat c.age_n i.illhealth i.deprivation i.employed i.region_n i.prov_id

scalar N_m=e(N)

///Loop over all estimated terms in e(b) and post coef/SE/pval for each level matrix b=e(b) local terms:
colnames b

foreach t of local terms { if "`t'"=="_cons" continue

///coef and SE for this specific level/term scalar b_m=_b[t] scalar se_m=_se[t]

///z and two-sided p-value scalar z_m=b_m/se_m scalar p_m=2*normal(-abs(z_m))

post P' ("y'") ("t'") (N_m) (b_m) (se_m) (z_m) (p_m) } }

postclose P' use miss_level_pvals', clear order outcome term N coef se z pval sort outcome term list,
sepy(outcome) clean

use `touse', clear

//===== //MAIN ANALYSIS //=====

tempfile modelres tempname M

postfile M' str40 model_id str32 outcome str12 model_type double N double treat_coef double treat_se
double treat_ci_lo /// double treat_ci_hi double p0 double p1 double cohen_h double h_lo double h_hi
double outcome_sd double hedges_g /// double g_lo double g_hi double pval using modelres', replace

///Cohen's h model-Primary Outcome (FSQ lvl 1) melogit outcome_n treat i.region_n i.female i.ethnicity2
age_n i.illhealth deprivation i.employed || prov_id: , vce(cluster prov_id)

scalar N_m=e(N) scalar b_m=_b[treat] scalar se_m=_se[treat] scalar z=b_m/se_m scalar pval=2normal(-
abs(z)) scalar lo_m=b_m-1.96se_m scalar hi_m=b_m+1.96*se_m

quietly margins, at(treat=(0 1)) predict(mu) matrix T=r(table) scalar p0_m=T[1,1] scalar p1_m=T[1,2]

scalar h_m=2asin(sqrt(p1_m))-2asin(sqrt(p0_m)) scalar se_hm=sqrt(2/N_m) scalar h_lo_m=h_m-
1.96se_hm scalar h_hi_m=h_m+1.96se_hm scalar list h_m h_lo_m h_hi_m

post `M' ("melogit_outcome_n") ("outcome_n") ("melogit") (N_m) (b_m) (se_m) (lo_m) (hi_m) (p0_m)
(p1_m) (h_m) (h_lo_m) (h_hi_m) (.) (.) (.) (.) (pval)

///Cohen's g model 1: Secondary Outcome (Maths Confidence) mixed svy_end_fsq_qconf_n treat
svy_bs_qconf_n i.region_n i.female i.ethnicity2 age_n i.illhealth deprivation i.employed /// || prov_id:,
vce(cluster prov_id)

scalar N1=e(N) scalar b1=_b[treat] scalar se1=_se[treat] scalar z=b1/se1 scalar pval=2normal(-abs(z))
scalar lo1=b1-1.96se1 scalar hi1=b1+1.96*se1

estat sd, post scalar sd_cluster=_b[prov_id:sd(_cons)] scalar sd_resid=_b[Residual:sd(e)] scalar
var_cluster=sd_cluster^2 scalar var_resid=sd_resid^2 scalar D_total=sqrt(var_cluster+var_resid) scalar
g_total=b1/D_total scalar se_g_total=se1/D_total scalar g_total_lo=g_total-1.96se_g_total scalar
g_total_hi=g_total+1.96se_g_total

```

```

post `M' ("mixed_svy_end_fsq_qconf_n") ("svy_end_fsq_qconf_n") ("mixed") (N1) (b1) (se1) (lo1) (hi1) (.)
(.) (.) (.) (.) (se_g_total) (g_total) (g_total_lo) (g_total_hi) (pval)

//Cohen's g model 2: Secondary Outcome (Maths Confidence-Composite) mixed confidence_maths_end
treat confidence_maths_bs i.region_n i.female i.ethnicity2 age_n i.illhealth deprivation i.employed /// ||
prov_id: , vce(cluster prov_id)

scalar N2=e(N) scalar b2=_b[treat] scalar se2=_se[treat] scalar z=b2/se2 scalar pval=2normal(-abs(z))
scalar lo2=b2-1.96se2 scalar hi2=b2+1.96*se2

estat sd, post scalar sd_cluster=_b[prov_id:sd(_cons)] scalar sd_resid=_b[Residual:sd(e)] scalar
var_cluster=sd_cluster^2 scalar var_resid=sd_resid^2 scalar D_total=sqrt(var_cluster+var_resid) scalar
g_total=b2/D_total scalar se_g_total=se2/D_total scalar g_total_lo=g_total-1.96se_g_total scalar
g_total_hi=g_total+1.96se_g_total

post `M' ("mixed_confidence_maths_end") ("confidence_maths_end") ("mixed") (N2) (b2) (se2) (lo2) (hi2) (.)
(.) (.) (.) (.) (se_g_total) (g_total) (g_total_lo) (g_total_hi) (pval)

postclose M' preserve use modelres', clear order model_id outcome model_type N treat_coef treat_se
treat_ci_lo treat_ci_hi p0 p1 cohen_h h_lo h_hi outcome_sd hedges_g g_lo g_hi sort model_type outcome

save "model_results.dta", replace restore

*===== //ANALYSIS: UNCONDITIONAL MODELS
*===== tempfile modelres tempname M postfile M' str40 model_id str32
outcome str12 model_type double N double treat_coef double treat_se double treat_ci_lo /// double
treat_ci_hi double p0 double p1 double cohen_h double h_lo double h_hi double outcome_sd double
hedges_g /// double g_lo double g_hi double pval using modelres', replace

//Step 1: Estimate empty models to get unconditional total variance//

//Empty model for svy_end_fsq_qconf_n mixed svy_end_fsq_qconf_n || prov_id:, vce(cluster prov_id) estat
sd, post scalar sd_cluster_empty1=_b[prov_id:sd(_cons)] scalar sd_resid_empty1=_b[Residual:sd(e)]
scalar D_total_empty1=sqrt(sd_cluster_empty1^2+sd_resid_empty1^2)

//Empty model for confidence_maths_end mixed confidence_maths_end || prov_id:, vce(cluster prov_id)
estat sd, post scalar sd_cluster_empty2=_b[prov_id:sd(_cons)] scalar sd_resid_empty2=_b[Residual:sd(e)]
scalar D_total_empty2=sqrt(sd_cluster_empty2^2+sd_resid_empty2^2)

//Step 2: Run full models and compute effect sizes//

//Cohen's h model-Primary Outcome (FSQ lvl 1) melogit outcome_n treat i.region_n i.female i.ethnicity2
age_n i.illhealth deprivation i.employed || prov_id: , vce(cluster prov_id) scalar N_m=e(N) scalar
b_m=_b[treat] scalar se_m=_se[treat] scalar z=b_m/se_m scalar pval=2normal(-abs(z)) scalar lo_m=b_m-
1.96se_m scalar hi_m=b_m+1.96se_m quietly margins, at(treat=(0 1)) predict(mu) matrix T=r(table) scalar
p0_m=T[1,1] scalar p1_m=T[1,2] scalar h_m=2asin(sqrt(p1_m))-2asin(sqrt(p0_m)) scalar
se_hm=sqrt(2/N_m) scalar h_lo_m=h_m-1.96se_hm scalar h_hi_m=h_m+1.96*se_hm scalar list h_m
h_lo_m h_hi_m post `M' ("melogit_outcome_n") ("outcome_n") ("melogit") (N_m) (b_m) (se_m) (lo_m)
(hi_m) (p0_m) (p1_m) (h_m) (h_lo_m) (h_hi_m) (.) (.) (.) (.) (pval)

//Cohen's g model 1: Secondary Outcome (Maths Confidence)-using empty model variance mixed
svy_end_fsq_qconf_n treat svy_bs_qconf_n i.region_n i.female i.ethnicity2 age_n i.illhealth deprivation
i.employed /// || prov_id:, vce(cluster prov_id) scalar N1=e(N) scalar b1=_b[treat] scalar se1=_se[treat]
scalar z=b1/se1 scalar pval=2normal(-abs(z)) scalar lo1=b1-1.96se1 scalar hi1=b1+1.96se1 scalar

```

```

g_total=b1/D_total_empty1 scalar se_g_total=se1/D_total_empty1 scalar g_total_lo=g_total-1.96se_g_total
scalar g_total_hi=g_total+1.96*se_g_total post `M' ("mixed_svy_end_fsq_qconf_n")
("svy_end_fsq_qconf_n") ("mixed") (N1) (b1) (se1) (lo1) (hi1) (.) (.) (.) (.) (.) (D_total_empty1) (g_total)
(g_total_lo) (g_total_hi) (pval)

```

```

//Cohen's g model 2: Secondary Outcome (Maths Confidence-Composite)-using empty model variance
mixed confidence_maths_end treat confidence_maths_bs i.region_n i.female i.ethnicity2 age_n i.illhealth
deprivation i.employed /// || prov_id: , vce(cluster prov_id) scalar N2=e(N) scalar b2=_b[treat] scalar
se2=_se[treat] scalar z=b2/se2 scalar pval=2normal(-abs(z)) scalar lo2=b2-1.96se2 scalar hi2=b2+1.96se2
scalar g_total=b2/D_total_empty2 scalar se_g_total=se2/D_total_empty2 scalar g_total_lo=g_total-
1.96se_g_total scalar g_total_hi=g_total+1.96*se_g_total post `M' ("mixed_confidence_maths_end")
("confidence_maths_end") ("mixed") (N2) (b2) (se2) (lo2) (hi2) (.) (.) (.) (.) (.) (D_total_empty2) (g_total)
(g_total_lo) (g_total_hi) (pval)

```

```

postclose M' preserve use modelres', clear order model_id outcome model_type N treat_coef treat_se
treat_ci_lo treat_ci_hi p0 p1 cohen_h h_lo h_hi outcome_sd hedges_g g_lo g_hi sort model_type outcome
save "model_results_uncond.dta", replace restore

```

```

*===== //ICC ESTIMATION *=====

```

```

//UNCONDITIONAL MODELS//

```

```

tempfile modelres tempname M postfile M' str40 model_id str32 outcome str12 model_type double N
double treat_coef double treat_se double treat_ci_lo /// double treat_ci_hi double p0 double p1 double
cohen_h double h_lo double h_hi double outcome_sd double hedges_g /// double g_lo double g_hi double
pval using modelres', replace

```

```

// Empty model for svy_end_fsq_qconf_n mixed svy_end_fsq_qconf_n || prov_id:, vce(cluster prov_id) estat
sd, post scalar sd_cluster_empty1=_b[prov_id:sd(_cons)] scalar sd_resid_empty1=_b[Residual:sd(e)]
scalar D_total_empty1=sqrt(sd_cluster_empty1^2+sd_resid_empty1^2)

```

```

// Empty model for confidence_maths_end mixed confidence_maths_end || prov_id:, vce(cluster prov_id)
estat sd, post scalar sd_cluster_empty2=_b[prov_id:sd(_cons)] scalar sd_resid_empty2=_b[Residual:sd(e)]
scalar D_total_empty2=sqrt(sd_cluster_empty2^2+sd_resid_empty2^2)

```

```

//Run full models and compute effect sizes// //Cohen's h model-Primary Outcome (FSQ lvi 1) melogit
outcome_n treat i.region_n i.female i.ethnicity2 age_n i.illhealth deprivation i.employed || prov_id: ,
vce(cluster prov_id) scalar N_m=e(N) scalar b_m=_b[treat] scalar se_m=_se[treat] scalar z=b_m/se_m
scalar pval=2normal(-abs(z)) scalar lo_m=b_m-1.96se_m scalar hi_m=b_m+1.96se_m quietly margins,
at(treat=(0 1)) predict(mu) matrix T=r(table) scalar p0_m=T[1,1] scalar p1_m=T[1,2] scalar
h_m=2asin(sqrt(p1_m))-2asin(sqrt(p0_m)) scalar se_hm=sqrt(2/N_m) scalar h_lo_m=h_m-1.96se_hm
scalar h_hi_m=h_m+1.96*se_hm scalar list h_m h_lo_m h_hi_m post `M' ("melogit_outcome_n")
("outcome_n") ("melogit") (N_m) (b_m) (se_m) (lo_m) (hi_m) (p0_m) (p1_m) (h_m) (h_lo_m) (h_hi_m) (.) (.)
(.) (.) (pval)

```

```

//Cohen's g model 1: Secondary Outcome (Maths Confidence)-using empty model variance mixed
svy_end_fsq_qconf_n treat svy_bs_qconf_n i.region_n i.female i.ethnicity2 age_n i.illhealth deprivation
i.employed /// || prov_id:, vce(cluster prov_id) scalar N1=e(N) scalar b1=_b[treat] scalar se1=_se[treat]
scalar z=b1/se1 scalar pval=2normal(-abs(z)) scalar lo1=b1-1.96se1 scalar hi1=b1+1.96se1 scalar
g_total=b1/D_total_empty1 scalar se_g_total=se1/D_total_empty1 scalar g_total_lo=g_total-1.96se_g_total
scalar g_total_hi=g_total+1.96*se_g_total post `M' ("mixed_svy_end_fsq_qconf_n")
("svy_end_fsq_qconf_n") ("mixed") (N1) (b1) (se1) (lo1) (hi1) (.) (.) (.) (.) (.) (D_total_empty1) (g_total)
(g_total_lo) (g_total_hi) (pval)

```

```
//Cohen's g model 2: Secondary Outcome (Maths Confidence-Composite)-using empty model variance
mixed confidence_maths_end treat confidence_maths_bs i.region_n i.female i.ethnicity2 age_n i.illhealth
deprivation i.employed /// || prov_id: , vce(cluster prov_id) scalar N2=e(N) scalar b2=_b[treat] scalar
se2=_se[treat] scalar z=b2/se2 scalar pval=2normal(-abs(z)) scalar lo2=b2-1.96se2 scalar hi2=b2+1.96se2
scalar g_total=b2/D_total_empty2 scalar se_g_total=se2/D_total_empty2 scalar g_total_lo=g_total-
1.96se_g_total scalar g_total_hi=g_total+1.96*se_g_total post `M' ("mixed_confidence_maths_end")
("confidence_maths_end") ("mixed") (N2) (b2) (se2) (lo2) (hi2) (.) (.) (.) (.) (.) (D_total_empty2) (g_total)
(g_total_lo) (g_total_hi) (pval)
```

```
postclose M' preserve use modelres', clear order model_id outcome model_type N treat_coef treat_se
treat_ci_lo treat_ci_hi p0 p1 cohen_h h_lo h_hi outcome_sd hedges_g g_lo g_hi sort model_type outcome
save "model_results_uncond.dta", replace restore
```

```
//UNCONDITIONAL MODELS// tempfile modelres tempname M postfile M' str40 model_id str32 outcome
str12 model_type double N double treat_coef double treat_se double treat_ci_lo /// double treat_ci_hi
double p0 double p1 double cohen_h double h_lo double h_hi double outcome_sd double hedges_g ///
double g_lo double g_hi double pval using modelres', replace
```

```
// Empty model for svy_end_fsq_qconf_n mixed svy_end_fsq_qconf_n || prov_id:, vce(cluster prov_id) estat
sd, post scalar sd_cluster_empty1=_b[prov_id:sd(_cons)] scalar sd_resid_empty1=_b[Residual:sd(e)]
scalar D_total_empty1=sqrt(sd_cluster_empty1^2+sd_resid_empty1^2)
```

```
// Empty model for confidence_maths_end mixed confidence_maths_end || prov_id:, vce(cluster prov_id)
estat sd, post scalar sd_cluster_empty2=_b[prov_id:sd(_cons)] scalar sd_resid_empty2=_b[Residual:sd(e)]
scalar D_total_empty2=sqrt(sd_cluster_empty2^2+sd_resid_empty2^2)
```

```
//Run full models and compute effect sizes// //Cohen's h model-Primary Outcome (FSQ lvl 1) melogit
outcome_n treat i.region_n i.female i.ethnicity2 age_n i.illhealth deprivation i.employed || prov_id: ,
vce(cluster prov_id) scalar N_m=e(N) scalar b_m=_b[treat] scalar se_m=_se[treat] scalar z=b_m/se_m
scalar pval=2normal(-abs(z)) scalar lo_m=b_m-1.96se_m scalar hi_m=b_m+1.96se_m quietly margins,
at(treat=(0 1)) predict(mu) matrix T=r(table) scalar p0_m=T[1,1] scalar p1_m=T[1,2] scalar
h_m=2asin(sqrt(p1_m))-2asin(sqrt(p0_m)) scalar se_hm=sqrt(2/N_m) scalar h_lo_m=h_m-1.96se_hm
scalar h_hi_m=h_m+1.96*se_hm scalar list h_m h_lo_m h_hi_m post `M' ("melogit_outcome_n")
("outcome_n") ("melogit") (N_m) (b_m) (se_m) (lo_m) (hi_m) (p0_m) (p1_m) (h_m) (h_lo_m) (h_hi_m) (.) (.)
(.) (.) (pval)
```

```
//Cohen's g model 1: Secondary Outcome (Maths Confidence)-using empty model variance mixed
svy_end_fsq_qconf_n treat svy_bs_qconf_n i.region_n i.female i.ethnicity2 age_n i.illhealth deprivation
i.employed /// || prov_id:, vce(cluster prov_id) scalar N1=e(N) scalar b1=_b[treat] scalar se1=_se[treat]
scalar z=b1/se1 scalar pval=2normal(-abs(z)) scalar lo1=b1-1.96se1 scalar hi1=b1+1.96se1 scalar
g_total=b1/D_total_empty1 scalar se_g_total=se1/D_total_empty1 scalar g_total_lo=g_total-1.96se_g_total
scalar g_total_hi=g_total+1.96*se_g_total post `M' ("mixed_svy_end_fsq_qconf_n")
("svy_end_fsq_qconf_n") ("mixed") (N1) (b1) (se1) (lo1) (hi1) (.) (.) (.) (.) (.) (D_total_empty1) (g_total)
(g_total_lo) (g_total_hi) (pval)
```

```
//Cohen's g model 2: Secondary Outcome (Maths Confidence-Composite)-using empty model variance
mixed confidence_maths_end treat confidence_maths_bs i.region_n i.female i.ethnicity2 age_n i.illhealth
deprivation i.employed /// || prov_id: , vce(cluster prov_id) scalar N2=e(N) scalar b2=_b[treat] scalar
se2=_se[treat] scalar z=b2/se2 scalar pval=2normal(-abs(z)) scalar lo2=b2-1.96se2 scalar hi2=b2+1.96se2
scalar g_total=b2/D_total_empty2 scalar se_g_total=se2/D_total_empty2 scalar g_total_lo=g_total-
1.96se_g_total scalar g_total_hi=g_total+1.96*se_g_total post `M' ("mixed_confidence_maths_end")
("confidence_maths_end") ("mixed") (N2) (b2) (se2) (lo2) (hi2) (.) (.) (.) (.) (.) (D_total_empty2) (g_total)
(g_total_lo) (g_total_hi) (pval)
```

```
postclose M' preserve use modelres', clear order model_id outcome model_type N treat_coef treat_se
treat_ci_lo treat_ci_hi p0 p1 cohen_h h_lo h_hi outcome_sd hedges_g g_lo g_hi sort model_type outcome
save "model_results_uncond.dta", replace restore
```

```
*===== //SUB-GROUP ANALYSIS *===== tempfile modelres
tempname M
```

```
postfile M' str40 model_id str32 outcome str12 model_type double N /// double treat_coef double treat_se
double treat_ci_lo /// double treat_ci_hi double p0 double p1 double cohen_h double h_lo double h_hi ///
double outcome_sd double hedges_g double g_lo double g_hi double pval /// using modelres', replace
```

```
local re "" || prov_id:" local vceopt "vce(cluster prov_id)"
```

```
///Ethnicity/// melogit outcome_n i.treat###i.ethnicity2 i.region_n i.female age_n i.illhealth deprivation
i.employed /// re', vceopt'
```

```
scalar N_m=e(N)
```

```
///Get subgroup-specific p0 and p1 in a way that aligns by margins colnames quietly margins ethnicity2,
at(treat=0) predict(mu) matrix B0=r(b) local cols: colnames B0
```

```
quietly margins ethnicity2, at(treat=1) predict(mu) matrix B1=r(b)
```

```
foreach cn of local cols {
```

```
///Extract subgroup level k from margins column name local k=. if regexm("`cn'", "[0-9]+.") local
k=real(regexs(1))
```

```
scalar p0_m=B0[1,"cn"] scalar p1_m=B1[1,"cn"]
```

```
///Cohen's h and your simple CI scalar h_m=2*asin(sqrt(p1_m))-2*asin(sqrt(p0_m)) scalar
se_hm=sqrt(2/N_m) scalar h_lo_m=h_m-1.96*se_hm scalar h_hi_m=h_m+1.96*se_hm
```

```
///Interaction term for this level (base level has no interaction parameter) scalar b_int=0 scalar se_int=.
scalar lo_int=. scalar hi_int=. scalar p_int=.
```

```
if k'<. { capture confirm number k' if _rc==0 { capture scalar b_int=_b[1.treat#k'.ethnicity2] if _rc==0 { scalar
se_int=_se[1.treat#k'.ethnicity2] scalar lo_int=b_int-1.96*se_int scalar hi_int=b_int+1.96*se_int scalar
z_int=b_int/se_int scalar p_int=2*normal(-abs(z_int)) } } }
```

```
post M' ("melogit_outcome_n_ethnicity2_cn") ("outcome_n") ("melogit_ethnicity2_int") (N_m) (b_int)
(se_int) (lo_int) (hi_int) (p0_m) (p1_m) (h_m) (h_lo_m) (h_hi_m) /// (.) (.) (.) (.) (p_int) }
```

```
///Female/// melogit outcome_n i.treat###i.female i.region_n i.ethnicity2 age_n i.illhealth deprivation
i.employed re', vceopt'
```

```
scalar N_m=e(N)
```

```
quietly margins female, at(treat=0) predict(mu) matrix B0=r(b) local cols: colnames B0
```

```
quietly margins female, at(treat=1) predict(mu) matrix B1=r(b)
```

```
foreach cn of local cols {
```

```
local k=. if regexm("`cn'", "[0-9]+.") local k=real(regexs(1))
```

```

scalar p0_m=B0[1,"cn"] scalar p1_m=B1[1,"cn"]

scalar h_m=2asin(sqrt(p1_m))-2asin(sqrt(p0_m)) scalar se_hm=sqrt(2/N_m) scalar h_lo_m=h_m-
1.96se_hm scalar h_hi_m=h_m+1.96se_hm

scalar b_int=0 scalar se_int=. scalar lo_int=. scalar hi_int=. scalar p_int=.

if k'<. { capture scalar b_int=_b[1.treat#k'.female] if _rc==0 { scalar se_int=_se[1.treat#k'.female] scalar
lo_int=b_int-1.96se_int scalar hi_int=b_int+1.96se_int scalar z_int=b_int/se_int scalar p_int=2*normal(-
abs(z_int)) } }

post M' ("melogit_outcome_n_female_cn") ("outcome_n") ("melogit_female_int") (N_m) (b_int) (se_int)
(lo_int) (hi_int) (p0_m) (p1_m) (h_m) (h_lo_m) (h_hi_m) /// (. ) ( . ) ( . ) ( . ) (p_int) }

///ill-health/Disability melogit outcome_n i.treat##i.illhealth i.region_n i.female i.ethnicity2 age_n deprivation
i.employed /// re', vceopt'

scalar N_m=e(N)

quietly margins illhealth, at(treat=0) predict(mu) matrix B0=r(b) local cols: colnames B0

quietly margins illhealth, at(treat=1) predict(mu) matrix B1=r(b)

foreach cn of local cols {

local k=. if regexm("`cn`","([0-9]+).") local k=real(regexs(1))

scalar p0_m=B0[1,"cn"] scalar p1_m=B1[1,"cn"]

scalar h_m=2asin(sqrt(p1_m))-2asin(sqrt(p0_m)) scalar se_hm=sqrt(2/N_m) scalar h_lo_m=h_m-
1.96se_hm scalar h_hi_m=h_m+1.96se_hm

scalar b_int=0 scalar se_int=. scalar lo_int=. scalar hi_int=. scalar p_int=.

if k'<. { capture scalar b_int=_b[1.treat#k'.illhealth] if _rc==0 { scalar se_int=_se[1.treat#k'.illhealth] scalar
lo_int=b_int-1.96se_int scalar hi_int=b_int+1.96se_int scalar z_int=b_int/se_int scalar p_int=2*normal(-
abs(z_int)) } }

post M' ("melogit_outcome_n_illhealth_cn") ("outcome_n") ("melogit_illhealth_int") (N_m) (b_int) (se_int)
(lo_int) (hi_int) (p0_m) (p1_m) (h_m) (h_lo_m) (h_hi_m) /// (. ) ( . ) ( . ) ( . ) (p_int) }

postclose M' use modelres', clear

use `touse', clear

//===== //CACE ANALYSIS //===== gen itt_treat=. replace itt_treat=0 if treat==0
replace itt_treat=1 if tot_sess>=12 & tot_sess!=. //80% Compliance replace itt_treat=0 if tot_sess<12 &
tot_sess!=.

preserve tempname handle postfile `handle' session n_treated n_attend pct_attend /// using
treat_attendance_by_session, replace

foreach s of numlist 1/15 {

// treated with non-missing response count if treat== 1 & !missing(session`s') local N_treated=r(N)

```

```

// treated who attended all or part count if treat== 1 & /// inlist(session`s', "Attended all", "Attended part")
local N_attend=r(N)

local pct=100 * N_attend'/N_treated'

// post results post handle' (s') (N_treated') (N_attend') (`pct') }

postclose `handle' use treat_attendance_by_session, clear restore

local controls "i.region_n i.female i.ethnicity2 age_n i.illhealth deprivation i.employed"

tempname H tempfile iv_results

postfile H' str40 model str40 depvar double b se lci uci double N double fs_b fs_se fs_lci fs_uci double
fs_b_pct fs_lci_pct fs_uci_pct /// double fs_F fs_r2 using "iv_results", replace

///FSQ lvl 1/// ivregress 2sls outcome_n `controls' (itt_treat=treat), vce(cluster prov_id)

local b=_b[itt_treat] local se=_se[itt_treat] local N=e(N) local tcrit=invttail(e(df_r), 0.025) local lci=b'-tcrit'*se'
local uci=b'+tcrit'*se'

///Manual first stage quietly regress treat itt_treat controls', vce(cluster prov_id) local fs_b=_b[itt_treat] local
fs_se=_se[itt_treat] local fs_lci=fs_b'-tcrit'*fs_se' local fs_uci=fs_b'+tcrit'*fs_se' quietly test itt_treat local
fs_F=r(F) local fs_r2=e(r2)

post H' ("FSQ lvl 1") ("outcome_n") (b') (se') (lci') (uci') (N') (fs_b') (fs_se') (fs_lci') (fs_uci') (100*fs_b')
(100*fs_lci') (100*fs_uci') /// (fs_F') (`fs_r2')

///Math Confidence/// ivregress 2sls svy_end_fsq_qconf_n svy_bs_qconf_n `controls' (itt_treat=treat),
vce(cluster prov_id)

local b=_b[itt_treat] local se=_se[itt_treat] local N=e(N) local tcrit=invttail(e(df_r), 0.025) local lci=b'-tcrit'*se'
local uci=b'+tcrit'*se'

quietly regress treat itt_treat svy_bs_qconf_n controls', vce(cluster prov_id) local fs_b=_b[itt_treat] local
fs_se=_se[itt_treat] local fs_lci=fs_b'-tcrit'*fs_se' local fs_uci=fs_b'+tcrit'*fs_se' quietly test itt_treat local
fs_F=r(F) local fs_r2=e(r2)

post H' ("Confidence") ("svy_end_fsq_qconf_n") (b') (se') (lci') (uci') (N') (fs_b') (fs_se') (fs_lci') (fs_uci')
(100*fs_b') (100*fs_lci') (100*fs_uci') /// (fs_F') (`fs_r2')

///Math Confidence-composite/// ivregress 2sls confidence_maths_end confidence_maths_bs `controls'
(itt_treat=treat), vce(cluster prov_id)

local b=_b[itt_treat] local se=_se[itt_treat] local N=e(N) local tcrit=invttail(e(df_r), 0.025) local lci=b'-tcrit'*se'
local uci=b'+tcrit'*se'

quietly regress treat itt_treat confidence_maths_bs controls', vce(cluster prov_id) local fs_b=_b[itt_treat]
local fs_se=_se[itt_treat] local fs_lci=fs_b'-tcrit'*fs_se' local fs_uci=fs_b'+tcrit'*fs_se' quietly test itt_treat
local fs_F=r(F) local fs_r2=e(r2)

post H' ("Confidence composite") ("confidence_maths_end") (b') (se') (lci') (uci') (N') (fs_b') (fs_se') (fs_lci')
(fs_uci') (100*fs_b') (100*fs_lci') (100*fs_uci') /// (fs_F') (`fs_r2')

```

```

postclose H' preserve use "iv_results", clear order model depvar b se lci uci N fs_b fs_se fs_lci fs_uci
fs_b_pct fs_lci_pct fs_uci_pct fs_F fs_r2

save "iv_summary_results.dta", replace restore

tab treat, missing

//===== //MULTIPLE IMPUTATION (MICE)
//===== preserve keep if treat!=.

misstable summarize age_n female ethnicity2 illhealth employed deprivation svy_bs_qconf_n
confidence_maths_bs mi set wide mi register imputed deprivation ethnicity2 employed svy_bs_qconf_n
confidence_maths_bs mi register regular age_n treat prov_id region_n female illhealth

mi impute chained (regress) svy_bs_qconf_n confidence_maths_bs (logit, augment) deprivation ethnicity2
employed= female illhealth age_n treat i.prov_id i.region_n, add(10) rseed(123) force

mi estimate, post: mixed outcome_n treat i.region_n i.female i.ethnicity2 age_n i.illhealth i.deprivation
i.employed || prov_id:, vce(cluster prov_id)

restore

local all outcome_n confidence_maths_bs ethnicity2 employed svy_bs_qconf_n foreach v of local all { gen
miss_v'=missing(v') }

mean miss_*, over(treat)

local prov "prov_id" local treat "treat"

local xlsx "outcomes_effectsizes.xlsx" capture erase ""xlsx"

///Continuous outcomes local y_cont_end "svy_end_fsq_qconf_n confidence_maths_end " local
y_cont_base "svy_bs_qconf_n confidence_maths_bs"

///Binary outcomes local y_bin_end "outcome_n" local y_bin_base "."

///Covariates used in outcome models local X "i.region_n i.female i.ethnicity2 c.age_n i.illhealth
c.deprivation i.employed"

local kc1: word count y_cont_end' local kc2: word count y_cont_base'

local kb1: word count y_bin_end' local kb2: word count y_bin_base'

tempfile touse save `touse', replace

preserve rename (svy_end_fsq_qconf_n confidence_maths_end svy_bs_qconf_n confidence_maths_bs
treat) (conf2 conf_comp2 conf1 conf_comp1 treat1) gen treat2=treat1 keep treat1 treat2 conf2 conf_comp2
conf1 conf_comp1 uniquelearnerid reshape long conf conf_comp treat, i(uniquelearnerid) j(time) hist conf,
by(treat time) color("16 79 117") graphregion(color(white)) plotregion(color(white)) xtitle("Confidence in
Maths") bin(5) //hist conf_comp, by(time) restore

tempfile raw postfile hraw str40 outcome str10 wave double N_all mean_all sd_all double N_c mean_c
sd_c double N_t mean_t sd_t using `raw', replace

forvalues i=1/kc1' { local y: word i' of y_cont_end' local y0: word i' of `y_cont_base'

```

```

///endline quietly count if !missing(y') & !missing(treat') local Nall=r(N) quietly summarize y' if !missing(y') &
!missing(`treat') local mall=r(mean) local sdall=r(sd)

quietly count if treat'==0 & !missing(y') local N0=r(N) quietly summarize y' if treat'==0 & !missing(`y') local
m0=r(mean) local sd0=r(sd)

quietly count if treat'==1 & !missing(y') local N1=r(N) quietly summarize y' if treat'==1 & !missing(`y') local
m1=r(mean) local sd1=r(sd)

post hraw ("y") ("end") (Nall') (mall') (sdall') (N0') (m0') (sd0') (N1') (m1') (sd1')

///baseline (if available) if "y0" != "." { quietly count if !missing(y0') & !missing(treat') local Nall=r(N) quietly
summarize y0' if !missing(y0') & !missing(treat') local mall=r(mean) local sdall=r(sd)

quietly count if treat'==0 & !missing(y0') local N0=r(N) quietly summarize y0' if treat'==0 & !missing(`y0')
local m0=r(mean) local sd0=r(sd)

quietly count if treat'==1 & !missing(y0') local N1=r(N) quietly summarize y0' if treat'==1 & !missing(`y0')
local m1=r(mean) local sd1=r(sd)

post hraw ("y") ("base") (Nall') (mall') (sdall') (N0') (m0') (sd0') (N1') (m1') (sd1') } }

///binary outcomes (means are proportions) for values i=1/kb1' { local y: word i' of y_bin_end' local y0: word
i' of `y_bin_base'

///endline quietly count if !missing(y') & !missing(treat') local Nall=r(N) quietly summarize y' if !missing(y') &
!missing(`treat') local mall=r(mean) local sdall=r(sd)

quietly count if treat'==0 & !missing(y') local N0=r(N) quietly summarize y' if treat'==0 & !missing(`y') local
m0=r(mean) local sd0=r(sd)

quietly count if treat'==1 & !missing(y') local N1=r(N) quietly summarize y' if treat'==1 & !missing(`y') local
m1=r(mean) local sd1=r(sd)

post hraw ("y") ("end") (Nall') (mall') (sdall') (N0') (m0') (sd0') (N1') (m1') (sd1')

///baseline if "y0" != "." { quietly count if !missing(y0') & !missing(treat') local Nall=r(N) quietly summarize y0'
if !missing(y0') & !missing(treat') local mall=r(mean) local sdall=r(sd)

quietly count if treat'==0 & !missing(y0') local N0=r(N) quietly summarize y0' if treat'==0 & !missing(`y0')
local m0=r(mean) local sd0=r(sd)

quietly count if treat'==1 & !missing(y0') local N1=r(N) quietly summarize y0' if treat'==1 & !missing(`y0')
local m1=r(mean) local sd1=r(sd)

post hraw ("y") ("base") (Nall') (mall') (sdall') (N0') (m0') (sd0') (N1') (m1') (sd1') } } postclose hraw

preserve use raw', clear format mean_all sd_all mean_c sd_c mean_t sd_t %9.3f format N_all N_c N_t
%9.0f export excel using "xlsx", sheet("raw_summary") firstrow(variables) sheetreplace restore

use `touse', clear

///Continuous endline outcomes foreach v of local y_cont_end {

```

```

local xt "Score" if strpos("v","comp") | strpos("v","Comp") | strpos("v","composite") |
  strpos("v","Composite") { local xt "Composite measure" } else if strpos("v","confidence") |
  strpos("v","Confidence") { local xt "Single measure" }

capture graph drop h_v' histogram v' if !missing(v'), name(h_v', replace) color(navy) lcolor(navy)
graphregion(color(white)) plotregion(color(white)) xtitle("`xt'")

graph export "hist_`v'.png", replace }

///Binary endpoint outcomes foreach v of local y_bin_end { capture graph drop h_v' histogram v' if
!missing(v'), name(h_v', replace) discrete color(navy) lcolor(navy) graphregion(color(white))
plotregion(color(white)) xtitle("v'") graph export "hist_`v'.png", replace }

use `touse', clear

///Continuous outcomes: tempfile effc postfile hec str40 outcome double b se ci_lo ci_hi p N double
var_prov var_resid sd_total icc_prov double d_model d_lo d_hi using `effc', replace

forvalues i=1/kc1' { local y: word i' of y_cont_end' local y0: word i' of `y_cont_base'

if "y0"=="." { quietly mixed y' c.treat' X' || prov':, vce(cluster prov') } else { quietly mixed y' c.treat' c.y0' X' ||
prov':, vce(cluster prov') }

local N=e(N) local b=_b[treat'] local se=_se[treat'] local z=b/se' local p=2normal(-abs(z')) local ci_lo=b'-
1.96se' local ci_hi=b'+1.96*`se'

quietly estat sd matrix Vc=r(Var) local var_prov=Vc[1,1] local var_resid=Vc[2,1]

local sd_total=sqrt(var_prov+var_resid') local icc=var_prov/(var_prov+`var_resid')

local d=b/sd_total' local d_lo=ci_lo/sd_total' local d_hi=ci_hi/sd_total'

post hec ("y'") (b') (se') (ci_lo') (ci_hi') (p') (N') (var_prov') (var_resid') (sd_total') (icc') (d') (d_lo') (d_hi') }
postclose hec

preserve use effc', clear format b se ci_lo ci_hi p var_prov var_resid sd_total icc_prov d_model d_lo d_hi
%9.3f format N %9.0f export excel using "xlsx", sheet("effects_cont") firstrow(variables) sheetreplace
restore

///Binary outcomes: melogit tempfile effb postfile heb str40 outcome double logOR se ci_lo ci_hi p N double
OR OR_lo OR_hi double var_prov icc_prov double d_approx d_lo d_hi using `effb', replace

forvalues i=1/kb1' { local y: word i' of y_bin_end' local y0: word i' of `y_bin_base'

if "y0"=="." { quietly melogit y' c.treat' X' || prov':, vce(cluster prov') } else { quietly melogit y' c.treat' c.y0' X'
|| prov':, vce(cluster prov') }

local N=e(N) local b=_b[treat'] local se=_se[treat'] local z=b/se' local p=2normal(-abs(z')) local ci_lo=b'-
1.96se' local ci_hi=b'+1.96*`se'

local OR=exp(b') local OR_lo=exp(ci_lo') local OR_hi=exp(`ci_hi')

local var_prov=_b[/var(_cons[prov'])] local icc=var_prov/(`var_prov'+(_pi^2/3))

local d=b'/1.81 local d_lo=ci_lo'/1.81 local d_hi=`ci_hi'/1.81

```

```
post heb ("y") (b) (se) (ci_lo) (ci_hi) (p) (N) (OR) (OR_lo) (OR_hi) (var_prov) (icc) (d) (d_lo) (d_hi) }
postclose heb
```

```
preserve use effb', clear format logOR se ci_lo ci_hi p OR OR_lo OR_hi var_prov icc_prov d_approx d_lo
d_hi %9.3f format N %9.0f export excel using "xlsx", sheet("effects_bin") firstrow(variables) sheetreplace
restore
```

```
use `touse', clear
```

```
local all svy_bs_qnumconf_aq2_n svy_bs_qnumconf_bq2_n svy_bs_qnumconf_cq2_n
svy_bs_qnumconf_dq2_n svy_bs_qnumconf_eq2_n svy_bs_qnumconf_fq2_n local all_end
svy_end_fsq_qnumconf_aq2_n svy_end_fsq_qnumconf_bq2_n svy_end_fsq_qnumconf_cq2_n
svy_end_fsq_qnumconf_dq2_n svy_end_fsq_qnumconf_eq2_n svy_end_fsq_qnumconf_fq2_n
```

```
tokenize all' tempfile results tempname post postfile post' str40 outcome str40 baseline double b se ll ul p N
using `results', replace
```

```
local k=0 foreach v of local all_end { local k=`k'+1 local base=""`k"
```

```
quietly mixed v' treat base' i.region_n i.female i.ethnicity2 age_n i.illhealth deprivation i.employed || prov_id:
, vce(cluster prov_id)
```

```
quietly lincom treat local b=r(estimate) local se=r(se) local ll=r(lb) local ul=r(ub) local p=r(p) local N=e(N)
```

```
post post' ("v") ("base") (b) (se) (ll) (ul) (p) (N) } postclose post'
```

```
use `results', clear format b se ll ul %9.4f format p %9.4g list, sepby(outcome)
```

Appendix 6: Additional tables

Table 29: Attrition predictiveness

Term	Confidence in maths (composite) – Estimate	Confidence in maths (composite) – P-value	Confidence in maths – Estimate	Confidence in maths – P-value	Attainment – Estimate	Attainment – P-value
Treatment (ref = Control)	1.59	0.183	1.60	0.180	-	-
Female (ref = Male)	0.34	0.150	0.34	0.150	0.61	0.684
Mixed/ Multiple ethnic groups	-0.32	0.525	-0.32	0.526	-	-
Asian or Asian British	-0.05	0.898	-0.14	0.694	2.53	0.062
Black, Black British, Caribbean or African	-0.45	0.144	-0.44	0.137	-	-
Other ethnic group (ref = White)	-0.53	0.256	-0.50	0.286	-	-
Age	-0.003	0.715	-0.004	0.652	0.005	0.937
Ill health (ref = Not ill health)	-0.26	0.235	-0.24	0.278	0.45	0.715
Deprived (ref = Not deprived)	0.10	0.667	0.15	0.527	0.81	0.576
Employed (ref = Not employed)	-0.17	0.444	-0.11	0.602	-0.22	0.877
East of England	1.80	0.206	1.79	0.209	-1.83	0.260
London	-0.34	0.741	-0.30	0.765	-	-

Term	Confidence in maths (composite) – Estimate	Confidence in maths (composite) – P-value	Confidence in maths – Estimate	Confidence in maths – P-value	Attainment – Estimate	Attainment – P-value
North East	-	-	-	-	-	-
North West	-0.91	0.403	-1.55	0.175	-	-
South East	1.71	0.111	1.67	0.119	-	-
Yorkshire and the Humber (ref = East Midlands)	-0.14	0.896	-0.16	0.877	-	-
Provider ID	2.33	0.138	2.94	0.068	-	-
Provider ID	1.47	0.304	1.47	0.305	-	-
Provider ID	-0.04	0.902	-0.03	0.914	-	-
Provider ID	1.59	0.221	1.58	0.223	-	-
Provider ID	-	-	-	-	-	-
Provider ID	-	-	-	-	-	-
Provider ID	-2.49	0.051	-2.45	0.054	-0.67	0.661
Provider ID	1.07	0.478	1.08	0.474	-	-
Provider ID	1.67	0.062	1.62	0.070	-	-
Provider ID	3.75	0.000	3.71	0.000	-	-
Provider ID	-0.02	0.988	-0.008	0.995	-	-
Provider ID	-	-	-	-	-	-
Provider ID	-	-	-	-	-	-
Provider ID	0.08	0.878	0.09	0.859	-	-
Provider ID	-0.82	0.529	-0.84	0.523	-	-
Provider ID	-	-	-	-	-	-
Provider ID	2.36	0.110	2.37	0.108	-	-
Provider ID	-3.01	0.024	-2.99	0.025	-	-
Provider ID	-	-	-	-	-	-

Term	Confidence in maths (composite) – Estimate	Confidence in maths (composite) – P-value	Confidence in maths – Estimate	Confidence in maths – P-value	Attainment – Estimate	Attainment – P-value
Provider ID	-	-	-	-	-	-
Provider ID	-	-	-	-	-	-
Provider ID	-	-	-	-	-	-
Provider ID	-	-	-	-	-	-
Provider ID	-	-	-	-	-	-
Provider ID	-	-	-	-	-	-

Source: ILR.

Base: n = 914.

Note: Dashes indicate that a certain variable was dropped by the analysis model if that variable/category had no association with missingness (i.e. if within a certain variable/category all values were present).



Department
for Education

© Department for Education copyright 2026

This publication is licensed under the terms of the Open Government Licence v3.0, except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3.

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

Reference: RR1628/7

ISBN: 978-1-83870-795-8

For any enquiries regarding this publication, contact www.gov.uk/contact-dfe.

This document is available for download at www.gov.uk/government/publications.