Department for Education

Adult numeracy randomised controlled trials: Contextualised approach

Cluster-randomised pilot impact and implementation and process

May 2026

Becci Newton, Seemanti Ghosh, Susie Bamford, Nick Litsardopoulos, Amy Fox, Kyla Ellis, Olivia Garner, IES

Government Social Research

Contents

List of tables

List of figures

Acknowledgements

The authors would like to thank the staff from adult education institutions who supported the trial, participated in interviews and shared their experiences of delivering the Contextualised Approach (CA). We are sincerely grateful to the learners who completed surveys and/or contributed through interviews, providing valuable insights to support the research.

We also wish to express our thanks to colleagues at the Department for Education (DfE) for their constructive feedback on the research tools and report and for their support throughout the design and evaluation stages. We extend our appreciation to the managed service supplier (MSS), Eito, and the product developers, the Education and Training Foundation (ETF), for their participation in interviews, intervention development, marketing and trial recruitment.

Our thanks go to the research team at IES for their thoughtful contributions to the fieldwork, analysis and reporting: Laura Dewar, Beatrice Rosolin, Kyla Ellis, Alexandra Nancarrow and Olivia Garner. We are grateful to Sara Butcher and Ehecatl Hunt Duarte for their careful proofreading of the final report. Our sincere thanks go to Ipsos UK for the collection of data for the trial through surveys and the Ipsos Data Portal. We thank our consortium leads, Susan Mackay and Naomi Williamson from Ipsos UK for their support throughout delivery and our consortium partners at RAND Europe, Learning and Work Institute, and Kings College London.

Executive summary

This report sets out findings from a pilot randomised controlled trial (RCT) of the Contextualised Approach (CA) for Functional Skills Qualification (FSQ) Level 1 maths. This ran as part of a 3-arm trial with the Adapted Mastery Approach (AMA) to FSQ Level 1 maths. Both were part of the Adult Numeracy Trials that were delivered as part of the government's Multiply programme which ran from 2022 to March 2025. They aimed to test whether these approaches could improve adult learners' maths attainment and confidence. The trial compared these interventions to a control group, where providers delivered their usual FSQ Level 1 maths courses without change or adaption. As a pilot, the trial provided a small-scale investigation into the feasibility of implementing a trial in this setting, but it was not powered to detect a small effect size.

Key findings

The main goal (or primary outcome) for the impact evaluation was to improve maths attainment as measured by the pass rate for FSQ Level 1 maths qualifications. The trial found no statistically significant evidence at the 95% confidence level[1] that the CA improved pass rates for FSQ Level 1 maths qualifications in the treatment group compared to the control group. While the treatment group saw a 1 percentage point (ppt) lower pass rate than the control group. However, this difference is not statistically significant, meaning it is not possible to be certain that it was caused by the intervention rather than by chance or a sampling error.

A further goal (or secondary outcome) for the impact evaluation was to improve maths confidence. The evaluation found no statistically significant evidence at the 95% confidence level that the CA improved learners' overall confidence in maths in the treatment group compared to the control group. These maths confidence analyses were limited by low survey response rates (45% for treatment learners and 33% for control group learners). Two confidence measures were used: 'overall confidence level in maths at the end of the trial' and a composite of confidence scores across specific everyday financial scenarios requiring maths (e.g., understanding interest rates on bank statements). Notably, these 2 confidence measures produced estimates in opposite directions. The treatment group's overall confidence score was 0.17 points lower than the control group, while their composite

[1] Differences that are statistically significant at the 95% confidence level are greater than would be expected by chance alone. This means we can be 95% certain that the difference is real and not due to sampling error. The 95% confidence level is widely accepted as the standard threshold for determining statistical significance in research and evaluation.

confidence score was 0.24 points higher, confidence being measured on a Likert scale where 1 meant not confident at all and 5 meant very confident.

As would be expected since this trial was not powered to detect small effects, no evidence was found of different impacts across learner demographic characteristics, including sex, age, ethnicity, deprivation and health/disability status.

The qualitative evidence showed that both tutors who delivered the CA and learners who took part were very engaged with the intervention. Treatment learners expressed strong satisfaction with the programme, particularly the delivery style and real-life contexts which they felt increased their problem-solving capabilities. Learners who experienced the CA were more likely to report being “very confident” in maths than the control group, but when “fairly” and “very” confidence categories were combined, there was no discernible difference between the two groups.

In business as usual (BAU) practices, tutors typically explained maths theories and concepts before offering real-life contexts for applying them. By contrast, CA begins with real-world examples to provide learners with tangible applications for sense-making. As such there was no evidence of overlap between trial arms.

Two fifths (42%) of providers dropped out following randomisation (9 treatment, 12 control). However, tutors who delivered the intervention did so according to the intervention guidance. Evidence suggested that the CA is feasible to deliver with minor adaptations and development.

Implementation barriers included late notice of tutor training dates, the intensity of tutor training and incomplete coverage of the FSQ curriculum. Concerns were also raised about the intervention’s suitability for different learner groups, and these are areas requiring consideration for in any future roll out.

Overall, while the impact evaluation did not find statistically significant impacts, the pilot provides promising signs about the feasibility of a conducting a future RCT in adult learning settings. It demonstrates that providers and learners can be recruited successfully, that randomisation at the provider level is workable, and that outcome data can be collected.

Background and rationale

The Adult Numeracy Randomised Controlled 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 CA is underpinned by Realistic Mathematics Education (RME). This builds maths knowledge from the contexts in which maths problems occur to be meaningful to learners. Through this, learners come to understand both common maths concepts and their practical applicability in the world outside of the classroom. This helps them develop a deep and long-term understanding. RME requires a pedagogic practice and classroom culture where tutors take a facilitative role, enabling discussion among students. International evidence indicates the effectiveness of RME. There is also evidence from the UK that the RME approach improves problem-solving and increases confidence in maths among school pupils. However, it had not been tested with an adult population until this trial. The primary objective of this pilot RCT was to gather evidence on the effectiveness of this teaching approach on adult learners' maths attainment and confidence. A secondary objective was to generate learning about delivering RCTs in adult education settings.

Intervention

The RME approach was originally developed by a team at Manchester Metropolitan University (MMU) for use in schools at Key stage 2 and early Key stage 3. This was selected as the basis for CA because its curriculum broadly aligned with FSQ Level 1 maths requirements. FSQ Level 1 maths is a qualification that focuses on practical maths skills needed for everyday life and work, signalling to employers that learners possess foundational maths competencies.

The ETF adapted CA from MMU's school resources specifically for adult learners studying FSQ level 1 maths. The adapted lessons covered number, measure, shape and space and data, aligning with most adult Level 1 courses.

Tutors received 10 hours of initial training, structured lesson plans and teaching resources. They were expected to engage with 80% of the 2-hour continuing

professional development (CPD) sessions, which were delivered by an expert throughout the trial and scheduled weekly or fortnightly depending on how the intervention was delivered. For learners, the intervention comprised 18 hours of lessons over either 15 or 30 weeks, embedding the CA 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 the Contextualised Approach as part of their FSQ Level 1 Maths course, in comparison to adult learners who do not receive the Contextualised Approach 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 were obtained from the Individual Learner Record (ILR) [2] but were not available for all participating learners. Missing primary outcomes overwhelmingly correlated with learners not completing or dropping out of the FSQ Level 1 maths course. As documented in the statistical analysis plan, learners without FSQ test data were classified as having failed. This approach, combined with the lower than planned number of treatment providers and imbalanced numbers of learners across arms, likely reduced the estimated effects.

Due to challenges with data collection, maths confidence was the only secondary outcome included in the analysis. This was measured using a learner survey completed at the end of FSQ Level 1 maths courses. Two additional secondary outcomes (maths skills and attendance) had to be excluded due to challenges with collecting these data.

The impact evaluation was delivered as a 3-armed pilot RCT, with randomisation at the provider level. Results from the pilot trial's third arm, the Adapted Mastery Approach (AMA), are presented in a separate report (Vallis et al., 2026).

The pilot trial used a gateway design that merged 2 previously separate trials (CA and AMA) into a single study. Providers were initially offered the 3-arm trial (CA, AMA, and control). However, if a provider had previously engaged with one of the interventions or strongly opposed implementing one to the point of withdrawing, they were offered participation in a 2-armed trial excluding that intervention (e.g. CA and

[2] The ILR is an on-going collection of data about learners from training providers in the Further Education (FE) and Skills sector in England.

control, or AMA and control). Control providers delivered their usual FSQ Level 1 maths courses with no changes or adaptations.

The Managed Service Supplier (MSS) recruited adult learning providers from the 480 further education (FE) providers then delivering FSQ Level 1 maths courses. From these participating providers, learners aged 19 and over who were enrolled in a FSQ Level 1 maths qualification during the 2024/25 academic year were recruited. The recruitment target was 15 providers per arm without dropouts (known as attrition), with randomisation at the provider level. Ultimately, the CA arm included 14 providers and 204 learners across England, while the control arm included 15 providers and 515 learners.

As a pilot, this trial provided a small-scale investigation into the feasibility of trialling the intervention, but it was not powered to detect small effect sizes. The implementation and process evaluation (IPE) examined how the CA was implemented, including tutors' engagement with training and support, their adherence to the intervention guidance and their teaching experiences with CA. The IPE also explored learners' responses to the approach, their perceived outcomes, and barriers and enablers to delivery, to assess feasibility and readiness for wider rollout. The IPE explored participant experiences in both treatment and control groups through baseline and endline surveys, along with 103 semi-structured interviews with learners, tutors, and providers. The IPE also analysed monitoring information, including learner attendance and tutor session delivery data.

Following the trial feasibility assessment in June 2023, the trial protocol was initially registered in June 2024, with the second version registered in September 2024. Recruitment ran from May 2024 to February 2025, with randomisation taking place from June to August 2024. The intervention was delivered between September 2024 and July 2025, with baseline surveys running between September 2024 and March 2025. IPE fieldwork took place from October 2024 to April 2025, and endline surveys were conducted in June 2025. ILR outcome testing took place from January through March 2026.

Impact findings

Primary outcome

The primary outcome was measured using FSQ Level 1 maths pass rates. The analysis found no statistically significant evidence at the 95% confidence level ($p = 0.958$[3]) that the CA improves pass rates. Treatment learners had a 1 percentage

[3] The p-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 have occurred by chance.

point lower pass rate than the control group. However, this difference is not statistically significant, meaning it is not possible to be certain that it was caused by the intervention rather than by chance or a sampling error.

Secondary outcome

The secondary outcome was confidence in maths, measured through both overall maths confidence and a composite score of maths confidence based on various real-world maths-use scenarios. The analysis found no statistically significant evidence at the 95% confidence level that the CA improved learners' overall confidence in maths. Response rates limited the reliability of these findings, with 45% of treatment learners and 33% of control learners completing the survey. The 2 confidence measures produced estimates in opposite directions, and neither was statistically significant. The treatment group's overall confidence score was 0.17 points lower than the control group ($p = 0.223$), while their composite confidence score was 0.24 points higher ($p=0.197$).

Compliance-adjusted, missingness and subgroup analyses

When the analysis included only learners who met the attendance thresholds set by the product developer, the estimated impact on pass rates and robustness improved slightly, but results remained statistically insignificant at the 95% confidence level. Primary outcome data were missing for 38% of learners. Although this missingness was not predicted by treatment status or most baseline characteristics (except ethnicity and provider preference for CA in the trial gateway), it was strongly correlated with FSQ Level 1 maths course completion. When learners with missing primary outcomes were excluded, the treatment effect became weakly statistically significant (at the 90% confidence level[4] rather than the standard 95% level). However, treatment status marginally predicted completion status, but the treatment effect was not statistically significant among compliers.

Given the low sample sizes, the analysis found no evidence of different impacts across learner demographic characteristics, including in terms of sex, age, ethnicity, deprivation and health/disability status.

[4] Differences that are statistically significant at the 90% confidence level are greater than would be expected by chance alone, however weaker than the standard 95% confidence level. This means we can be 90% certain that the difference is real and not due to sampling error. The 90% confidence level is not widely accepted as the standard threshold for determining statistical significance in research and evaluation, the standard threshold is 95%.

Implementation and process findings

IPE findings indicate positive views of intervention delivery. Providers, tutors and learners all reported good experiences, with treatment learners reporting high levels of satisfaction. While control group tutors and learners described using everyday contexts, these did not form the starting point for learning as they did in the CA meaning there was no indication of overlap between treatment and control teaching. Most treatment learners appreciated the real-world approach, visual elements, group work and the variety of problem-solving methods. However, some treatment learners preferred the traditional approaches they experienced at school, particularly neurodivergent learners who found multiple solution methods confusing. Tutors saw the benefits of using real-world contexts before introducing maths concepts and appreciated both the CPD and their refreshed approach to teaching. Overall, the CA appeared feasible to deliver, with minor adaptation for learners requiring additional support.

It was relatively common for providers to withdraw following randomisation (known as attrition). Of the 50 providers recruited, 21 withdrew before or early in delivery. Among those providing reasons, withdrawal resulted from lack of buy-in from tutors, insufficient staff capacity for training or inability to recruit enough learners. In terms of engagement at the level defined by the product development (known as compliance), 10 of 13 tutors attended all the pre-training and over 80% of the CPD sessions.

Attendance data showed that 59% of learners attended 10 of the 12 sessions – the minimum level that product developers believed necessary for improved outcomes. Learners who attended fewer than 10 sessions cited personal and contextual barriers rather than course-related issues.

Product developers provided guidance on intervention delivery. The qualitative evidence suggested that tutors delivered the intervention with fidelity, meaning they followed the lesson plans and used materials as planned. However, some learners needed more support than the guidance suggested, taking up more time than the intervention allowed for. This led to some tutors offering additional catch-up sessions beyond the intervention model.

Overall, IPE evidence indicated that learners had a positive experience, but CA may need further adaptation to be fully inclusive and provide complete curriculum coverage. There is encouraging evidence that the intervention can be delivered in accordance with developer guidance. Providers and tutors endorsed the CA, with all saying they would recommend it to colleagues and other training providers. The ‘ready to go’ lesson materials and tools, including the ratio table and percentage bar,

were valued for their clarity and adaptability to other mathematics courses (pre-entry, level 2, etc).

Conclusions and recommendations

The impact evaluation found no statistically significant evidence that the CA improved pass rates or maths confidence among learners taking FSQ Level 1 maths courses. Nevertheless, the IPE established that the intervention was generally well received and largely delivered in line with developer guidance. Tutors welcomed the professional development and the refreshed approach to their teaching practice, while learners found the real-world application of maths engaging, noting it built their problem-solving capability and ability to understand maths in daily life.

However, several challenges may explain why these positive experiences did not translate into measurable impacts, offering lessons for the future. For example, multiple providers withdrew following randomisation and tutors faced barriers to completing the full training and CPD programme. This suggests potential practicality issues requiring reconsideration. Approximately 59% of learners completed the minimum number of sessions specified by the product developer. Non-attendance was typically due to personal barriers and may have affected pass rates regardless of the intervention. There was evidence that tutors adapted training and paced it differently than planned to accommodate some learner needs, although the extent of this was limited. These adaptations could be incorporated into future versions of CA.

International evidence indicates the effectiveness of RME, with UK studies also showing promise, though causal impact assessments are still pending. This pilot study provides valuable insights into RME's feasibility within a new setting and for a new learner group.

This study has substantial limitations, and findings should be interpreted accordingly. The impact evaluation was designed as a pilot to assess the feasibility of randomisation and CA delivery, but it was not powered to detect small effects. There were high rates of missing data for the primary outcome, correlating strongly with course completion status but not treatment allocation. The IPE modified data collection from the trial protocol due to practical constraints, research burden concerns, interview sampling limitations, and issues with learner survey response rates and missing data in administrative records. Nonetheless, these rich data provide insights into the diverse views and experiences of different participants.

Overall, while the impact evaluation did not find statistically significant impacts, the IPE evaluation provides promising signs about the feasibility of conducting RCTs in adult learning settings. It also demonstrates how the CA can be adapted to better accommodate diverse learner needs and tutor capacity.

However, this pilot reveals constraints for designing a fully powered trial of CA in adult FSQ Level 1 courses. The number of adult education providers nationally, at c500 in any year, is small compared to the number of providers in other phases of education. Although all eligible providers were approached, only around 10% participated, and even fewer completed delivery. A fully powered cluster-randomisation trial would require all eligible providers across England to sign up. Furthermore, pilot data on class numbers and sizes indicate that alternative randomisation strategies would not yield sufficient providers for a robust trial.

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).

1 Introduction

This report sets out findings from an impact evaluation and implementation and process evaluation (IPE) of the Contextualised Approach (CA) maths intervention, delivered in adult education settings. This novel intervention aimed to improve adult learners' maths attainment and confidence by teaching maths through the contexts in which it is used. The evaluation was part of a 3-armed, pilot randomised controlled trial (RCT), with CA as one treatment arm and the Adapted Mastery Approach (AMA) as the other. The 3rd arm, the 'control group', received business-as-usual (BAU) FSQ Level 1 maths delivery. The CA pilot included 719 learners across 29 providers. There were 515 learners in 15 control group providers, and 204 learners in 14 treatment providers. Results from the trial's second arm (AMA) are reported separately (Vallis et al., 2026).

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 declined over the decade prior to Multiply (DfE data). This trial sought to contribute evidence on using a CA to teach maths in adult education settings, specifically in Functional Skills Qualifications (FSQ) Level 1 maths classes.

The Realistic Mathematics Education (RME) approach underlying CA builds maths knowledge from contexts that are meaningful to learners. This starting point contrasts with traditional teaching methods, which typically introduce maths concepts and formulas before putting them in a context in which they are used.

A review by ETF identified several UK studies showing positive qualitative outcomes from Realistic Mathematics Education (RME), including improved problem solving and maths confidence among KS4 learners, with teachers reporting benefits (MMU, 2004–07; Durham University, 2010). More recent large-scale quantitative studies have produced mixed findings, partly due to COVID-19 disruptions and high attrition, particularly in post-16 trials (EEF, 2018; Hough et al., 2017; Zarnadze, 2021). No earlier trials were found involving adult FSQ Level 1 maths learners.

Earlier pilots, such as the Gatsby-funded MMU trial using MiC materials (2004–2007) and Durham University's 2010 re-evaluation, indicated better problem solving, strong conceptual understanding, and positive tutor and pupil feedback (Realistic Mathematics Education, n.d.b; Searle & Barmby, 2012). A small Irish study similarly found positive changes in classroom practice and learner engagement (Moffett & Corcoran, 2010). The EEF-funded MMU RCT (2018), involving 119 schools, found no measurable progress gains for RME students compared with controls and no impact for those on free school meals, though results were undermined by high attrition and COVID-19-related testing delays (Demack et al., 2022).

Post-16 studies show mixed but sometimes promising outcomes. The Nuffield-funded pilot in GCSE re-sit classes found no gains in algebra and unchanged attitudes, though small improvements emerged in one of two modules and qualitative analysis suggested improved discussion and meaning making (Hough et al., 2017). The Centre for Excellence in Maths (2021) reported that RME, particularly Ratio Tables, enhanced confidence, reasoning, and inclusivity, especially for lower-attaining learners (Zarnadze, n.d). A recent Greek quasi-experimental study found RME-based collaborative problem-solving improved performance and knowledge transfer, with skills retained a month after the intervention (Ventistas et al., 2024).

1.1.3 Rationale

Although some evidence exists on RME in school settings, the approach had not been trialled in England with adult learners undertaking FSQ Level 1 maths.

FSQ Level 1 maths was selected for this trial due to its accessibility to a broad range of adult learners and its value in developing maths confidence and skills applicable to everyday life and the workplace. The ETF, who designed the intervention, outlined the rationale for this choice in its summary report:

- Level 1 broadly aligns with the upper end of Key Stage 2 and early Key Stage 3 in schools, enabling ETF to build on and adapt resources that Manchester Metropolitan University (MMU) had developed using the RME approach to contextualisation.
- A Level 1 qualification in maths signals to employers that learners possess the foundational level of maths required for many roles.

1.2 Intervention description: theory of change

The CA intervention was designed to improve adult learners' confidence and achievement in Level 1 maths by adopting an RME approach.

The Theory of Change envisaged that financial resource and CA lesson resources (inputs) would enable recruitment of learners and tutors, and training of tutors and ongoing support (activities). It was expected that these activities would support delivery of 12 maths lessons centred around RME, exposing learners to real-world applications of maths as the vehicle for learning (outputs).

It was hoped these lessons would enable learners to connect real experiences to mathematical concepts, supporting better learning and retention of ideas. They would also understand the relevance of maths to their lives, increasing willingness to engage. In parallel, tutors would shift their approach from 'explaining' to 'eliciting' answers (change mechanisms). As a result, it was hoped that tutors would have improved subject and pedagogical knowledge, demonstrate improved listening skills create an inclusive RME classroom culture. Additionally, learners would be able to move easily between informal contexts and abstract mathematical representations, to mathematise and to engage more effectively with the FSQ curriculum (short-term outcomes). In the long-term, it was hoped that maths authority would shift from tutors to learners and that learners' maths capability and confidence would increase. Demonstrating this, learners would attempt questions in FSQ tests that they previously found hard (outcomes).

The ultimate intended impact was:

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

This Theory of Change is summarised in Table 1.

Table 1: Logic model for the Contextualised Approach (CA)

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 fell in the decade prior to Multiply.	**Aims**	CA aims to improve learner attainment and confidence in maths by using real-life application of maths as the vehicle for learning. It helps learners develop agency. Teaching methods focus on class discussions and experimentation with different problem-solving approaches, enabling learners to better understand the application of maths and become more open to maths learning.

Inputs and activities	Outputs	Change mechanisms	Outcomes	Impacts
Inputs • Financial resource • CA lesson resources **Activities** • Tutor recruitment • Learner recruitment • Tutor training • Ongoing CPD delivered by expert practitioners	• No. tutors recruited • No. training sessions delivered • No. learners recruited • Tutors deliver 12 lessons comprising teaching input, tools and opportunities based on the RME / CA	• Learners connect real experiences to maths ideas, and retain learning • Learners perceive maths as relevant to their lives, increasing willingness to engage and persist • Tutors shift their approach from 'explaining' to 'eliciting', giving learners more agency	**Short term** • Tutors improve listening and inclusion • Tutors create an RME classroom culture. • Learners gain improved understanding of mathematical concepts. • Learners move between informal contexts and abstract maths. • Learners can mathematise. • Learners engage more effectively with the FSQ curriculum. **Medium term** • Maths authority shifts from tutors to learners • Learners' maths capability increases • Learners' maths confidence increases • Learners attempt FS questions that they previously found harder relative to baseline	• Improved functional numeracy across adult population in England, with more attainment at level 1 FSQ • Long-term improvements in employability, confidence and independence • Social benefits, learners able to support children's learning and manage personal finances • Strengthened adult numeracy provision across the system

Evidence assessment	The intervention was informed by evidence of positive qualitative outcomes of the RME approach among key stage 4 students. Although results have been mixed, no studies were found trialling RME in England among adult learners undertaking FSQ Level 1 maths.

Assumptions	Tutors are given enough time to fully engage with the intervention at initial training and during ongoing weekly/fortnightly online CPD sessions	**Possible unintended consequences**	Learners could still experience maths anxiety.

1.3 Intervention Description[5]

Name

Contextualised Approach (CA) for FSQ Level 1 Maths

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

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 adult maths learning. A CA for FSQ Level 1 aimed to provide meaningful contexts through which learners would come to understand common maths concepts and the usefulness and applicability of maths for the world outside of the classroom.

See the sections on the intervention background and theory of change 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 functional skills tutors who were trained in CA. 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 structured lesson plans and supporting materials to aid the delivery of the CA. Lessons covered number, measure, shape and space, and data, to fit within most adult FSQ level 1 courses. Developed by ETF, lesson plans ensured all key topics were addressed clearly and effectively. More information on the intervention content can be found in Appendix 2.

Following completion of initial training, tutors were given access to all resources via a Padlet link[6]. Trainers also offered tutors regular online CPD support throughout the intervention delivery period, which tutors were expected to attend.

[5] 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.
[6] Padlet is an online collaboration, bulletin board app that is commonly used in education and training.

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

Tutors were expected to deliver 12 weekly 1.5-hour CA sessions to learners on 15-week courses. The product developers also provided a scheme that adapted pace to suit 30-week courses.

Who: Intervention providers/implementers

The product was developed by ETF and the RME team from Manchester Metropolitan University.

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

The intervention was delivered by FSQ Level 1 Maths tutors in providers that had been assigned to the CA treatment arm of the trial. Treatment tutors responsible for delivering the course were required to attend initial training and ongoing CPD sessions delivered by ETF. Training was mandatory and took place before the 2024/25 academic year, either in July or September 2024. This comprised 10 hours of online training across four 2.5-hour sessions. Tutors were also required to attend at least 80% of 12 weekly 2-hour online lesson study sessions, designed to support collaborative reflection, lesson planning, and pedagogical development.

How: Mode of delivery

The CA was delivered in the same mode as the FSQ maths Level 1 course teaching, which in practice was face-to face mode.

Where: Location of the intervention

Learners engaged with the intervention within adult education classes delivered by a trained tutor assigned to the treatment group. Providers were recruited from across England. The setting primarily involved adult education and FE colleges.

When and how much: duration and dosage of the intervention

The intervention was designed to be delivered in weekly 1.5-hour sessions over 12 weeks within 15-week courses. The maximum dosage was 18 hours. 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 24 weeks on 30-week courses.

Learners were expected to attend at least 80% of CA lessons, or around 14 hours.

The intervention started at the beginning of the 2024/25 academic year and continued until either Christmas 2024 or summer 2025, depending on each provider's timetable.

Some providers did not start delivery until early 2025, while others delivered the intervention to a second cohort of learners from that point. While ETF did not provide ongoing support to these later cohorts, they remained available for queries and ad hoc support. Data was collected on all learners, including those in later cohorts.

Tailoring: Adaption of the intervention

ETF supplied information on the degree of tailoring tutors could apply, such as varying contexts to better match learners' experience/recognition, but not the maths content. The intervention was developed with flexibility to accommodate diverse FE settings, although a minimum of 30 guided learning hours (GLH) was specified for a course to be included in the trial, to ensure sufficient time to account for individual learners' needs.

It was intended that the full content should be delivered without omission or significant modification. This was to ensure fidelity of the pedagogical approach while allowing for contextual responsiveness.

Modifications: Changes to the planned intervention

There were no major modifications to the planned intervention, but some tutors adapted practices to better accommodate learner needs.

Strategies to maximise effective implementation

Fidelity was assessed through the implementation and process evaluation (IPE), using quantitative surveys of learners 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 the IPE findings section.

1.4 Evaluation overview

This pilot RCT gathered evidence on the quality, relevance, suitability, and effectiveness of the CA course in adult education settings, specifically FSQ Level 1 maths courses. It explored the processes involved with implementing CA, and adaptations required to support future delivery or wider roll out. 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.

The focus of the impact evaluation was on testing whether the CA achieved its intended outcomes. It used an experimental design to assess the causal effect of using CA to teach maths in adult education settings on outcomes of interest. As part of a 3-arm gateway trial, which meant the control group could differ slightly from the other

intervention arm, CA was tested relative to BAU, i.e. the control group. Those providers allocated to the control condition delivered their usual FSQ Level 1 maths courses without change or adaptation.

For the IPE, evidence was gathered on the processes involved in delivery of the CA 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 Institute for Employment Studies (IES) 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 the Contextualised Approach as part of their FSQ Level 1 Maths course, in comparison to adult learners who do not receive the Contextualised Approach as part of their FSQ Level 1 Maths course?

Secondary research questions:

- What is the average difference in attendance, measured using records by providers, of adult learners in providers delivering the Contextualised Approach as part of their FSQ Level 1 Maths course, in comparison to adult learners who do not receive the Contextualised Approach in their FSQ Level 1 Maths course?
- What is the average difference in changes in confidence in maths, measured by pre-and-post-course surveys, of adult learners in providers who deliver Contextualised Approach as part of their FSQ Level 1 Maths course, in comparison to adult learners who do not receive the Contextualised Approach as part of their FSQ Level 1 Maths course?
- What is the average change in maths skills, measured using pre-post maths test results, of adult learners in providers delivering the Contextualised Approach as part of their FSQ Level 1 Maths course, in comparison to adult learners who do not receive the Contextualised Approach as part of their FSQ Level 1 Maths course?

1.5.2 Implementation and process evaluation

Main IPE research questions[7]:

[7] The full set of IPE research questions and corresponding sub-questions are detailed in Appendix 4

1. Has the intervention been delivered with fidelity (that is, in line with the intervention guidance)?
2. To what extent, if at all, do BAU FSQ Level 1 courses typically include any content relating to the CA?
3. Is there any evidence of contamination of the 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?
4. How did tutors experience delivering the intervention?
5. What was tutors' experience of the training and support provided to deliver the intervention?
6. What was learners' experience of the intervention?
7. What outcomes did the intervention have for learners?
8. What enablers and barriers were there to learner engagement and participation in the intervention?
9. What lessons have been learned for future delivery of the intervention or wider rollout?
10. What lessons have been learned from the pilot about delivering RCTs in adult education settings?

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 1Appendix 1: Glossary.

Further details about the costings, data collection tools and code used, and additional data tables and charts can be found in the appendices 2 –7.

2 Methods

2.1 Overview of trial design

Table 2 provides an overview of the trial design as outlined in the Pilot Trial Protocol[8]. The sections that follow provide further detail on the key design features.

Table 2: Overview of trial design

Trial element	Description
Trial name	Contextualised Approach (CA) for FSQ Level 1 Maths
Project title	A Realistic Maths Approach embedded into FSQ Level 1 maths courses
Developer	Education and Training Foundation (ETF)
Geography	England
Delivery	Adult learning providers
Evaluator	Institute for Employment Studies
Principal Investigators	Becci Newton, Seemanti Ghosh
Evaluation plan authors	Becci Newton, Seemanti Ghosh
Trial design	Cluster-randomised Pilot Gateway Trial with 3 arms; **First treatment arm:** Contextualised Approach intervention **Second treatment arm:** Adapted Mastery Approach intervention **Control arm:** FSQ Level 1 maths courses without either intervention
Trial type	Pilot gateway RCT
Trial population	Adults (19+) on a FSQ Level 1 maths course
Primary outcome	FSQ result (pass / fail)
Secondary outcomes	Attendance, Confidence in Maths, Maths Attainment

[8] The Pilot Trial Protocol can be accessed here: https://osf.io/46jrn/overview

Source: Pilot Trial protocol

2.2 Detailed trial design

This study was designed as a 3-armed parallel cluster randomised pilot trial, with CA as a treatment arm, AMA as a second treatment arm, and business-as-usual (BAU) as the control arm. Cluster randomisation occurred at the provider level, with stratification by region.

The primary outcome was the FSQ Level 1 maths pass rate, measured at the end of the academic year 2024/25. No baseline data were collected for the primary outcome. The primary analysis followed an intention-to-treat (ITT) approach, as specified in the trial protocol. As the primary outcome was binary, the estimated treatment effect is reported as a percentage point difference in pass rates between treatment and control groups.

Secondary outcomes included attendance, confidence in maths, and maths skills. Confidence in maths was assessed through baseline and endline learner surveys administered by Ipsos. The other secondary outcomes (attendance and maths skills) were excluded from the final analysis due to problems with data collection.

As a pilot RCT, this trial was only powered to detect very large effect sizes. The sample size for this trial was not expected to produce a reliable estimate of the impact of the intervention and its effect size.

2.2.1 Gateway Trials

This trial was constructed as a pilot trial by merging 2 previously planned trials, each originally designed to test 1 of the 2 interventions. This design change aimed to reduce the required sample size by using a single pooled control group for both interventions. However, there was a risk that providers might prefer 1 intervention over the other and withdraw rather than accept a 1-in-3 chance of receiving their non-preferred option. It was not possible to assess the likelihood of this before launching recruitment.

To address this, providers who declined the 3-armed pilot trial but would otherwise have withdrawn offered 1 of 2 alternative 2-armed trials, known as Gateway Trials. The MSS initially offered all providers the 3-armed trial, if providers strongly preferred a specific intervention, they were then offered the relevant 2-armed trial instead. When providers' tutors had already received AMA training (as AMA was being rolled out to learners aged 16-19 during the trial period), the MSS offered only gateway option 3, which focused on CA.

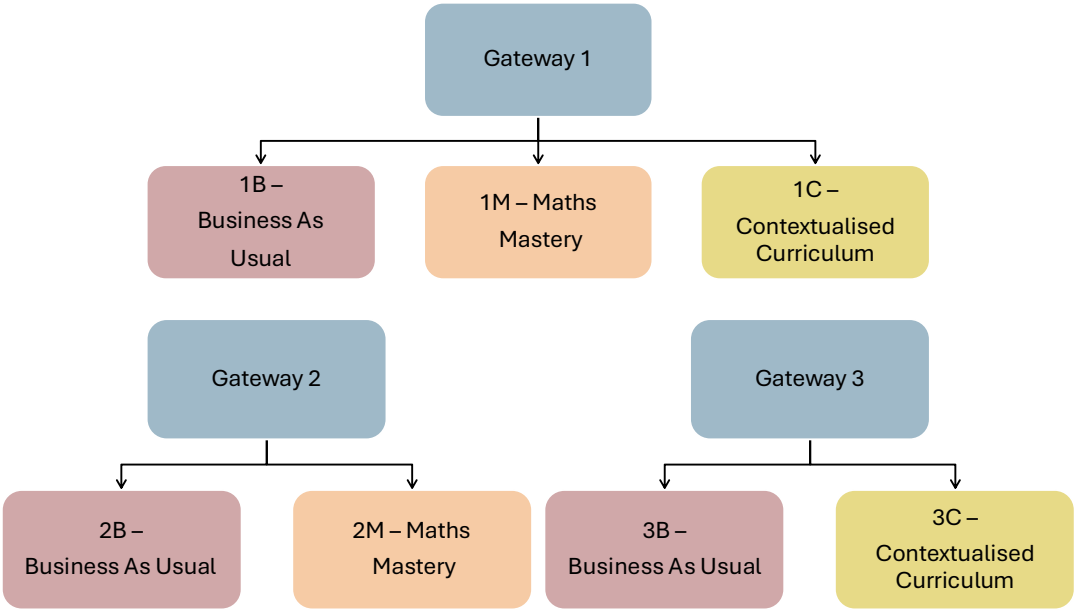
While providers entering Gateway Trials increased their probability of receiving their preferred intervention (from 33% to 50%), they simultaneously reduced their overall probability of receiving any intervention (from 66% to 50%). Given that both interventions were considered to be equally effective at the time, this trade-off offered no clear overall

advantage. It does, however, suggest a potential bias, intentional or otherwise, towards maximising the chance of receiving the preferred intervention, even at the cost of a higher likelihood of being assigned to the control condition. Each provider received the option that maximised expected value for their institution.

Under this design, the MSS first attempted to recruit all providers into the 3-armed pilot trial, represented as gateway 1 in Figure 1. If a provider expressed reservations or declined to participate due to the possibility of receiving either intervention, the MSS offered them gateway 2 (with a 50-50 chance of receiving AMA) or gateway 3 (with a 50-50 chance of receiving the CA). Through this choice, providers traded a higher chance of receiving any intervention for a higher chance of receiving their preferred one.

This design meant the control group was shared between AMA and CA in gateway 1, but not in gateways 2 or 3, where control groups were specific to each treatment. In practice, no providers chose gateway 2, while 12 providers chose gateway 3 (CA), mostly because of previous exposure to AMA-type interventions.

Figure 1: Gateway trial design



2.3 Recruitment

The MSS started recruitment to the trial in early June 2024, targeting the 480 adult learning providers in England delivering FSQ Level 1 maths courses. Participating providers then recruited tutors, who received information about the trial and could opt out.

Learners were also recruited through their learning providers. They received detailed trial information and could opt out at any point. Learners enrolled in courses during summer and autumn 2024. Those opting out could choose between opting out of primary data

collection only or to withdrawing from the pilot trial entirely. Learners who opted out of the trial were still permitted to access their courses, but were not invited to any further trial activities, and their data was not included in the trial.

2.3.1 Sample size

The original intention was to implement a full-scale trial of CA, but recruitment challenges meant this was not possible. This necessitated a revised approach of a multi-arm pilot Gateway Trial with AMA and a control arm. The pilot trial was conducted as a small-scale preliminary study to assess the feasibility of the interventions and running a trial in adult learning settings. Key components such as recruitment strategies and outcome data collection were tested, identifying potential barriers and facilitators to implementation, along with early indications of effect size. As an exploratory study, the pilot RCT was not statistically powered to detect treatment effects but helped establish assumptions and conditions for a fully powered trial (Knight, 2019). To ensure at least 30 clusters after accounting for potential attrition, the aim was to recruit 66 providers in total – 22 per trial arm. If all providers had opted for the gateway design, this would have resulted in 44 providers per individual pilot trial, or 30 after adjusting for up to 30% attrition.

The final CA analysis included 719 learners across 29 providers. There were 515 learners in 15 control group providers, and 204 learners in 14 treatment providers. Of the 29 providers, 7 opted for the CA Gateway meaning they were only part of CA sample. Table 3 presents the distribution of the sample size.

Table 3: Distribution of sample at analysis

Recruitment Gateway choice	Treatment Learners	Control Learners	Treatment Providers	Control Providers
3-arm	168	444	10	12
Gateway (2-arm)	36	71	4	3
Total	204	515	14	15

Note: Source: Management Information data. Base: All learners as randomised and analysed, 719

2.3.2 Randomisation

King's College London (KCL) generated the allocation for each provider at the point of recruitment (Appendix 5). Adult education providers were randomly assigned to treatment and control groups using a Stata-based algorithm designed to create 3 equally sized trial arms. Randomisation was stratified by region to ensure balanced representation.

As outlined earlier, some providers expressed interest in receiving one intervention but not the other, leading to the Gateway Trial option. When these Gateway Trials were

activated, KCL conducted randomisation using the same procedure but limited to 2 arms, either AMA versus Control or CA versus Control, instead of the standard 3-arm design.

Following randomisation, training sessions were scheduled for tutors at providers assigned to the treatment arms. Randomisation and training processes were conducted in 3 batches based on provider engagement and adherence to timelines. Training sessions were offered once in July and twice in September 2024. Learners were then automatically assigned to treatment or control status based on the allocation of their provider. Thus, any learner enrolled at a treatment-assigned provider became part of the treatment group, and likewise for control.

2.4 Outcome measures

2.4.1 Primary outcome measure

The primary outcome measure was the pass rate of FSQ Level 1 maths, sourced from the Individual Learner Record (ILR)[9]. The 'Outgrade' variable for each learner's FSQ Level 1 maths was used to determine pass/fail status. As learners can take the test multiple times throughout the year and typically stop after passing, the most recent grade obtained was used to assess achievement or failure. Where providers operated multiple cohorts under the 12-week delivery model, all cohorts were included in the analysis.

While the intervention was unlikely to affect participants' initial enrolment decisions, it might affect their likelihood of completing the course. To avoid confounding results, any participant who enrolled but subsequently dropped out or failed to complete their course was recorded as having failed (coded as 0 in the binary indicator using the ILR variable CompStatus).

2.4.2 Secondary outcome measures

Confidence in maths was the only secondary outcome include in the final analysis. The other 2 (maths skills and attendance) were excluded because endline maths skills tests could not be administered in most providers and attendance data was not consistently collected from providers.

Maths confidence was measured by baseline and endline learner surveys using a question from the national evaluation of Multiply. This Likert-type question asked about learners' overall level of confidence in maths ranging from 1 (not at all confident) to 5 (very confident). The specific question is shown in appendix 4 IPE survey, QCONF. A second set of questions explored confidence in everyday scenarios, using the same Likert scales, which was used to form the composite confidence maths score. Scenarios included checking change is correct when buying something, working out the best deals

[9] The ILR is an on-going collection of data about learners from training providers in the Further Education (FE) and Skills sector in England.

when shopping, helping children with maths homework, understanding interest rates on bank statements, keeping track of bank account balances and working with numbers as part of a job. The question is shown in full in appendix 4 IPE survey, question QNUMCONF.

2.5 Data collection

Data for the primary outcome, that is the test result for FSQ Level 1 maths, was collected from the ILR, which is the national-level dataset for recording further education learners and their outcomes in England. Data for the secondary outcome was collected using the learner survey (Appendix 4).

2.6 Data analysis

2.6.1 Primary analysis

The primary analysis for this pilot trial followed an intention-to-treat (ITT) approach. Given that the primary outcome was a binary variable, the estimated treatment effect reflected a percentage point change in the probability of treated individuals passing the FSQ Level 1 course compared to the control group receiving BAU as calculated in the Estimation of Effect Size section. To protect against potential bias, the analysts remained blinded to cohort assignments. The team achieved this by assigning data cleaning and analysis tasks to separate members. The analysis of the primary outcome used a multilevel logistic model of the form:

$$Y_{ij} = \Lambda \left(\beta_0 + \beta_1 T_j^M + \beta_2 T_j^C + \beta_3 X_{ij} + \beta_4 S_j + u_j \right)$$

Where between cluster variance is defined as:

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

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

2.6.2 Secondary analysis

Analysis of the secondary outcome (confidence in maths and composite confidence) used a baseline-adjusted linear mixed-effects model (ANCOVA). Endline confidence was regressed on the treatment indicator, the baseline confidence score, and region fixed

effects, with a random intercept for provider to account for clustering. The model used the following specification:

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

Where the cluster and residual variance terms are defined as:

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

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

Where all variables are defined as for the primary analysis and $Y_{t-1,ij}$ is the outcome at baseline. Standard errors clustered at the provider level.

2.6.3 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 18 hours of the intervention for various reasons, but for the compliance analysis they were considered compliant if they attended 10 or more of the 12 classes.

To model compliance, a binary variable was created: 1 if the individual met the compliance criteria and 0 otherwise. The initial step examined the level of non-compliance through descriptive statistics. In the protocol, it was stated that if over 80% of learners were deemed to have complied, it would not be necessary to conduct further compliance analysis. However, compliance was found to be below 80%, which meant estimating a specific Complier Average Causal Effect (CACE) using the instrumental variable (IV) approach with the 2-stage least squares method (Angrist et al., 1996). In the first stage, treatment assignment was used to predict actual treatment receipt to check how strongly randomisation influences uptake. In the second stage, the predicted (instrumented) treatment receipt was regressed on the pass/fail outcome, adjusting for baseline characteristics and using provider-clustered standard errors. The second-stage coefficient represents the causal effect of treatment among compliers, whilst removing bias from non-random patterns of treatment uptake. Given the smaller sample size of this pilot trial, the CACE estimate should be treated as exploratory. Additionally, it was assumed that there was no contamination between the treatment and control by virtue of the randomisation strategy at the cluster level, i.e. all learners within a provider are either assigned to treatment or control with minimal chances of interaction between both groups.

2.6.4 Missing data analysis

The analysis used administrative data sources for the primary outcome and covariates, so it was less susceptible to the impact of missing data than an outcome sourced from survey data. Where missing data exceeded 5% and was not deemed to be Missing

Completely at Random (MCAR) following assessment of the data pattern, auxiliary variables were used to determine whether the missing data could be Missing at Random (MAR). To do this, binary indicators were created to identify missing/non-missing (coded as 1/0), and logistic regression models were run 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).

2.6.5 Sub-group analysis

Exploratory subgroup analyses were conducted on the primary and secondary outcomes to assess whether the impact of the intervention varied across the following sub-groups:

- Sex: Baseline ILR data using the binary variable SEX.
- Age: Baseline ILR data using the variable DATEOFBIRTH to generate a continuous measure.
- Learning difficulty/disability or health issue: Baseline ILR data using the categorical variable LLDDHEALTHPROB.
- Ethnicity: Baseline ILR data using the categorical variable ETHNICITY.

2.6.6 Estimation of effect size

As the primary outcome was the pass rate for FSQ Level 1 maths, historical data was used to shape the assumptions for the analysis. 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 the formula:

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

Following the logistic regression, the respective proportions were calculated as the probability of passing the FSQ, estimated from the main model as the average predicted probability if everyone had been untreated:

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

And the average predicted probability if everyone were treated:

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

Alternatively, the observed probability of passing could be used, but observed probabilities mix the effect of treatment assignment with other factors specific to who happened to be in the control group (covariates, cluster composition, missingness patterns). In contrast, the model computes the marginal standardised (counterfactual)

probabilities. Using these probabilities, the upper and lower bounds of Cohen's h confidence interval were 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.

2.6.7 Additional analysis and robustness checks

In addition to the ITT, further analysis was conducted to assess how the estimated effect changed when excluding those whose outcome was recorded as a fail (0) due to not sitting the exam. This was not mentioned in the protocol but was identified at the analysis stage due to the high proportion of missing data for the primary outcome data.

2.6.8 Variations from protocol

The key variation was that secondary outcomes maths skills and attendance were excluded from analysis and reporting. Furthermore, due to high proportion of missingness in the primary outcome, an additional robustness check was considered where estimates were adjusted only for those with a recorded outcome.

2.6.9 Limitations

The key limitation was the high proportion of missingness in the primary outcome that increased uncertainty of the impact estimates and attribution of any effect to treatment. Although missingness in primary outcome was not predicted by treatment status, missingness was correlated with course completion status of FSQ Level 1 Maths course.

2.7 Implementation and process evaluation

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

- Provide evidence on how the intervention was delivered and integrated into the existing FSQ Level 1 maths curriculum, the extent to which it was delivered consistently across providers and tutors, as specified, and how it differed from usual practice.

- Explore tutor and learner experiences, and views on what worked well and less well in delivering the intervention, to identify lessons for potential future roll-out.
- Explore additional outcomes for learners and tutors that were anticipated to result from the intervention, but which were not a focus of the RCT.
- Investigate how outcomes and impacts emerged, and were perceived, including unanticipated outcomes and any differences the intervention made to learners' and tutors' experience and practice.
- Gather insights into the feasibility and acceptability of conducting RCTs in adult education settings, including enablers and barriers to participation, levels of attrition, compliance with data collection requirements, and experiences of tutors and learners in both treatment and control groups.

The IPE used mixed methods, collecting data from learners, tutors, and providers:

- Baseline and endline surveys of learners in both treatment and control groups.
- Qualitative interviews with learners in both treatment and control groups.
- Qualitative interviews with tutors in both treatment and control groups.
- Qualitative interviews with representatives of participating providers in both treatment and control groups.
- Attendance data for tutors' training sessions and learners' course attendance.

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

2.7.1 Data collection

Learner surveys with treatment and control

The baseline survey was administered to all learners. It included questions specifically related to the IPE, including learners' motivation to take part in the course, and their perceived outcomes related to attitude towards maths and willingness to use maths. The survey was delivered through 2 methods: online via an open link and computer-assisted telephone interviewing (CATI). Open links to the online survey were included in the Trial Readiness Packs distributed to providers by the MSS, who in turn shared these with learners. The Trial Readiness Packs also informed learners that they might be called by Ipsos to take part in the survey by telephone. This way, CATI fieldwork supplemented the online link to reach non-responders. The start and end dates were aligned with 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, whilst the endline survey was open through various waves from 27 November 2024 to 11 September 2025.

One to 2 weeks after the open link was expected to be delivered to learners by providers, CATI fieldwork commenced. CATI interviews included screening questions to check whether learners had already completed the survey online. If a learner had already completed online, then the CATI interview was terminated. Most baseline survey responses were completed online, suggesting that the delivery mode and distribution plan were appropriate. It also suggested that providers were successfully distributing the open link to learners in classrooms. The proportion of completed surveys via each mode was broadly in line with expectations.

The endline survey was administered to all participants following the same approach as the baseline survey. The endline survey had additional questions specifically related to the impact evaluation and IPE, including overall satisfaction with the course, perceived outcomes related to attitudes, changes to maths confidence, perception of how the course helped, and suggestions to improve the course. The FSQ trial endline survey was initially administered as an online survey accessed via an open link. All learners were given the same link, which was shared in the Trial Readiness Packs. The Trial Readiness Packs also informed learners that they might be called by Ipsos to take part in the survey by telephone. As with the baseline, CATI fieldwork supplemented the online link to reach non-responders. CATI was implemented for the baseline and then continued for the endline.

Once sufficient learner contact details were shared by providers via the Ipsos Data Portal, the survey was expanded to include additional delivery methods. Unique links to an email survey were sent directly by Ipsos to participants, and a new CATI campaign began with this sample, enabling consistency in data collection with the other trials. This was a strategic decision taken to align this survey with the Preparation for GCSE and Family Numeracy Trials. The survey content remained the same across all delivery modes. The original open link accessed through the web link and QR code in the Trial Readiness Packs remained available throughout the 30-week courses, allowing participants to continue to submit surveys via these links, and enabling providers to encourage learners to take part.

Achieved sample

The total numbers of survey responses are presented in Table 4 by treatment and control group. The average response rate was 33% at baseline and 52% at endline[10]. The lowest response rate at baseline was among the control group learners, at 28%. In both surveys, the response rate was lower in the control group than in the treatment group. The same learners did not necessarily respond to both surveys, which means

[10] 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 CA 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.

comparisons between baseline and endline should be treated with caution. Overall, 74 learners in the control group and 44 learners in the treatment group completed both surveys.

Table 4: Overview of number of responses and response rates for learner surveys

Survey (group)	Number of responses	Response rates %	Total learners
Baseline (treatment)	80	39%	204
Baseline (control)	146	28%	515
Endline (treatment)	117	57%	204
Endline (control)	237	46%	515

Note: Source: Ipsos learner surveys. Base: All baseline and endline survey respondents

Learner interviews: treatment and control

Learners were contacted near the end of their Level 1 FSQ maths course or shortly after completion, with interviews conducted between January and July 2025. A purposive sampling approach was used to recruit a diverse group of learners, drawing on data provided to Ipsos by providers via the Ipsos Data Portal upload facility for the Multiply RCTs. Whilst baseline survey data were unavailable, efforts were made to ensure diversity across gender, age, attendance, class size, and provider. The aim was to conduct 40 learner interviews in total: 30 with CA treatment group learners and 10 with control group learners. In practice, 32 interviews were conducted with learners from 12 CA providers, plus 10 interviews with control group learners. See Appendix 8 **Error! Reference source not found.** for further details of the sample breakdown.

Treatment group interviews explored learners' motivations for studying FSQ Level 1 maths, their experiences of the course and teaching methods (e.g. real-life examples, group work, problem-solving strategies), and perceived outcomes such as improved skills, confidence, and motivation. Barriers and enablers to participation, including trial-related experiences, were also discussed. A similar approach was used with the control group, with 10 interviews across providers. These interviews covered comparable themes, including course experience, teaching methods, outcomes, and factors influencing participation in further maths learning.

Interviews took place by telephone or video call and lasted 20 to 45 minutes. Learners received a £30 shopping voucher as a thank you for their time.

Tutor interviews: treatment and control

Depth interviews were conducted with 11 tutors from the CA treatment arm and 10 tutors from the control arm. The target was to interview around a third of tutors involved, which was exceeded. Treatment tutors were interviewed at 2 points: first after completing pre-delivery training and initial lessons (October–November 2024), and again after course

completion but before exam results (February–July 2025), with timing aligned to course end dates. Across the 14 providers in the treatment group, 8 tutors were interviewed in both points, 3 at the first point only, and 4 at the second point only. The 10 control group tutors were interviewed once, after their courses had ended. All tutors were contacted and invited to take part in interviews.

Tutor interviews explored CPD training, delivery experiences, perceived learner impacts, and challenges, as well as typical teaching practices and the trial's influence on their work. All interviews were conducted by phone or video, lasting around 45 minutes. No incentives were offered.

Stakeholder interviews

Depth interviews were conducted with 4 national-level stakeholders involved in the trial: the managed service supplier, and representatives of the product developers who delivered the CA tutor training. These interviews gathered feedback on their roles, experiences, and lessons learned from the trial.

Provider Interviews: treatment and control

Additional interviews were conducted with senior representatives from providers in both the CA and control groups. These individuals were responsible for deciding whether their organisation would participate, liaising with the MSS, and ensuring compliance with contractual requirements. Given the small sample sizes in both groups, all eligible providers were contacted to ensure comprehensive coverage of relevant perspectives. In total, representatives from 10 providers were interviewed from the treatment group and 10 from the control group.

Management information

Learner attendance information was collected by Ipsos from CA providers through the Ipsos Data Portal to measure how much of the intervention learners received (dosage). The MSS and product developer (ETF) collected and shared information about tutors' attendance at the CA training and lesson study sessions.

Sample sizes

Qualitative interviews with tutors, providers and learners

Table 5 gives details of the samples of learners, tutors and providers who participated in the IPE data collection.

Table 5: Overview of IPE interviews

Research participant type	Treatment	Control
Learners	32	10
Wave 1 tutors	11	[x]
Wave 2 tutors	12	10
Providers and stakeholders	14	10

Note: Source: Management Information data. Base: Interview participants

Demographic details for the treatment group learners (Table A7 - 1 Appendix 7) are:

- 20 women (63%) and 6 men (19%) were interviewed as well as 6 learners whose gender was not recorded (19%). The aim was to achieve a balanced split by gender, but challenges were faced in recruiting men in the learner group.
- In terms of age bands, 19% of those interviewed were aged 18-24 (6 learners), 62% were aged 25-49 (20 learners, and 19% were aged 50 and over (6 learners).

The treatment group learners were drawn from 12 different providers, with an average of 3 learners interviewed per provider. The sample covered 8 of the 9 regions in England.

2.8 Variations from protocol

2.8.1 Observations

The timing of the CA training sessions and the small group format meant that the evaluation team could not observe tutor training and lesson study sessions. This limited opportunities for direct understanding of training delivery approaches.

2.8.2 Endline tutor survey

An endline tutor survey was originally planned to collect views on tutors' prior use of CA in their teaching, when the trial was expected to be larger. However, the reduced number of tutors in the pilot trial meant they were all approached for interviews instead of completing a survey.

2.9 Limitations

The IPE comprised predominantly qualitative research with participants who responded to requests to join interviews. Qualitative research aims to establish range and diversity of views and cannot provide insight into the scale of these. It cannot be known how representative these are of all participants' views. In some areas, the IPE reported data from the Ipsos Data Portal and from the MSS. The former relied upon providers completing records in full and not all did this. The MSS data was recorded contemporaneously to recruitment of providers and contains summary information which could vary by provider. The IPE also drew upon the learner surveys. These are affected by low response rates and may be biased towards learners who were more engaged with their courses or who found the demands of their courses easier to manage.

2.10 Analysis

The survey data was analysed using descriptive statistics, namely frequencies and cross-tabulations where applicable.

Survey responses were matched to ILR and learner data collected 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, 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[11]. Finally, given the exploratory nature and modest sample sizes, weighting could have introduced additional uncertainty without meaningful accuracy gains. The survey results presented in the IPE findings should therefore be considered indicative rather than definitive.

All interviews were digitally recorded with the participants' consent and transcribed verbatim. The data were analysed using an Excel-based framework. Relevant information from interview transcripts was summarised under key themes, with illustrative quotes selected. This allowed for comparison and contrast of data across participants and themes. Emerging themes from the qualitative data were then compared with survey responses and training attendance data, to explore the extent to which perceptions were reflected across the wider sample.

[11] 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.

The synthesis was grounded in the Theory of Change. Each element, including assumptions, enablers, and barriers, was examined to assess whether it had been realised. Evidence from the RCT was used to evaluate the impact of the intervention on key outcomes, whilst qualitative and quantitative data from the learner survey and interviews were analysed to identify additional outcomes. Qualitative insights from tutors and learners were used to explore the mechanisms outlined in the Theory of Change, as well as the associated enablers, barriers, and assumptions.

3 Impact evaluation results

The pilot trial found no evidence that CA improved learners' probability of passing the FSQ Level 1 maths exam. The unadjusted comparison of pass rates showed a small difference favouring the treatment group (54% vs 49%), but this simple comparison does not account for clustering, covariates, or missing data. The difference disappeared once these factors were included in the formal impact models. The primary ITT analysis, estimated via multilevel logistic regression with provider random intercepts and region fixed effects, produced a treatment coefficient very close to zero. This was not statistically significant, with a confidence interval indicating high uncertainty around the effect. Provider-level clustering was substantial, indicating that provider differences played a strong role in learner attainment, but adjusting for this did not change the null treatment effect. Missing outcome data was substantial (38%) and strongly linked to course completion rather than treatment assignment. Importantly, treatment did not predict missingness, meaning there was no differential attrition bias. In robustness checks, when all learners with missing outcomes were dropped, the treatment effect was positive and marginally significant. However, there was suggestive evidence of treatment increasing course completion, and once completion was accounted for, the effect on passing among those who complied remained non-significant. Overall, the treatment did not improve FSQ Level 1 maths attainment. The estimated effects were small, imprecise, and consistently non-significant across ITT, imputed, compliance-adjusted, and sensitivity analyses.

3.1 Participants and attrition

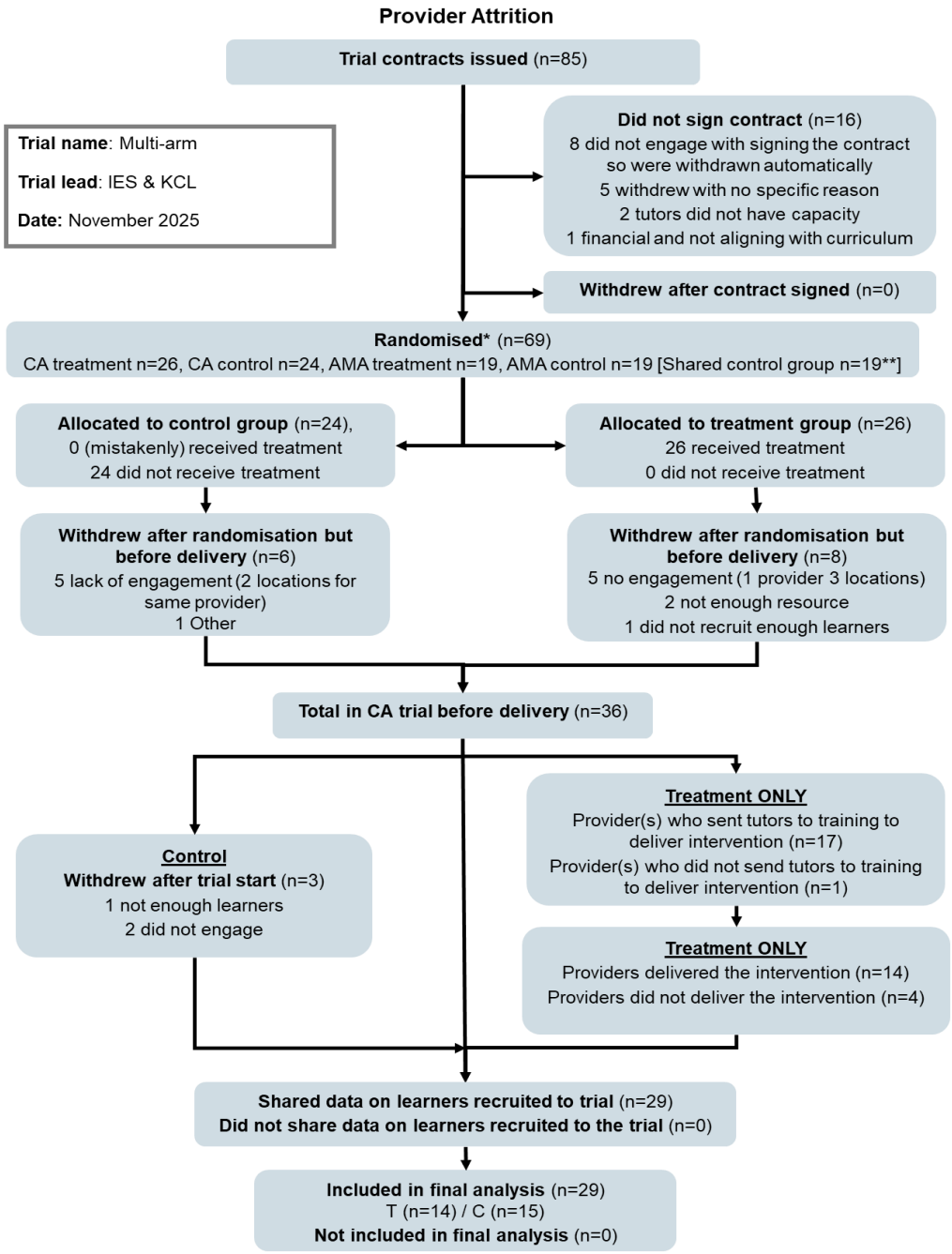
A total of 50 providers were recruited and randomised to the CA trial (24 to the control group and 26 to the treatment group). Eleven of these opted for the CA Gateway, which meant they could only receive CA or control. Of the 50 providers, 21 providers (~42%) withdrew after randomisation but before submitting any learner data (9 from the control group and 12 from the CA treatment arm), leaving a total sample size of 29 providers for analysis. Of these 29, 7 had opted for the CA Gateway Trial. Overall, reasons for withdrawal included lack of engagement, insufficient resource or not enough learners recruited. Trial attrition is shown in more detail in the CONSORT flow diagram in Figure 2 below.[12]

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. Table 6 summarises the distribution of providers across regions by treatment allocation at randomisation.

[12] 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.

There were 719 learners across these 29 providers; 515 learners in 15 control group providers, and 204 learners in 14 treatment providers. Table 6 shows the distribution of these 29 providers by region. The treatment and control groups both included providers from multiple regions, providing broad geographic coverage consistent with regional stratification used 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.

Figure 2: Provider attrition map



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

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

Table 6: Provider characteristics

Region	Treatment (At randomisation)	Treatment (At analysis)	Control (At randomisation)	Control (At analysis)
East Midlands	2	1	4	2
East of England	3	1	4	3
London	8	5	7	3
North East	1	1	1	1
North West	4	2	3	3
South East	2	2	1	1
South West	2	1	1	0
West Midlands	1	0	1	0
Yorkshire and the Humber	3	1	2	2
Total	26	14	24	15

Note: Source: Management information data. Base: All providers at randomisation and analysis

The primary outcome was pass or fail in FSQ Level 1 maths, derived from the ILR records. Since the primary outcome comes from administrative records, missing data on the primary outcome cannot be attributed directly to individual learners withdrawing and could be caused by issues at the provider level. However, 272 out of 719 learners (~38%) did not have a registered pass/fail status in the ILR. Out of these 272 learners, 65 learners (~24%) were not registered for the FSQ qualification aim, 122 (~45%) did not complete their course or withdrew and 85 learners (~31%) completed learning activities but did not have a pass/fail result registered.

The secondary outcome, confidence in maths, was collected using a survey. Overall, 195 learners responded to these questions at baseline, and 260 learners responded at endline. Survey responses were missing at baseline (73%) and endline (64%) for a large proportion of the sample. At baseline, among those who provided data, most participants reported being fairly or very confident, suggesting that less confident learners may have chosen not to respond altogether. Treatment assignment was broadly balanced across both arms. Table 7 shows the level of attrition from baseline for each of the confidence in maths measures. For the main confidence measure, only 82 out of the 195 baseline learners also responded at endline, representing an attrition rate of approximately 58%. However, attrition differed between the trial arms, with higher attrition in the control group (62%) compared to the treatment group (51%). The composite confidence measure showed a nearly identical pattern. The different attrition rates between groups suggests

that dropout patterns were likely to be systematically biased and related to treatment assignment.

Additionally, the overall level of attrition was high and substantially reduced the effective sample sizes for the secondary outcomes. This limited the study's ability to detect effects and increased uncertainty around the estimates. Further exploration of why data was missing and the use of multiple imputation to address this is described in Section 3.4.

Table 7: Attrition – secondary outcomes

Outcome	N (Baseline)	N (Endline)	Attrition rate	Attrition rate (Treatment)	Attrition rate (Control)
Confidence in maths	195	82	58%	51%	62%
Confidence in maths (composite)	194	81	58%	52%	62%

Note: Source: Baseline and endline survey. Base: All baseline and endline survey respondents

Table A6 - 1 in Appendix 6 summarises observed baseline characteristics. Missing data was less than 5% across demographic variables (sex, ethnicity and age), socio-economic indicators (deprivation and employment) and health status. The sample was primarily composed of individuals classified as having low deprivation (65%), with approximately one-third in the high deprivation category. High deprivation was defined as learners living in the bottom quintile (the 20% most deprived areas). Around 48% of participants were employed at baseline, and 48% were unemployed. The sample was ethnically diverse, with White participants representing majority of the sample (53%), substantial representation from Black (17%) and Asian (12%) ethnic groups. Females made up 79% of the sample, and around 28% of learners reported an illness or disability at baseline. Overall, the summary statistics suggest broadly comparable baseline characteristics across groups, whilst highlighting substantial missing data in the survey-based secondary outcome variables.

Table A6 - 2 in Appendix 6 shows the results of the baseline balance checks across key characteristics. Baseline balance was assessed by testing whether treatment assignment was related to individual-level baseline characteristics, controlling for region in all cases and accounting for provider clustering. The results indicate statistically significant differences between the treatment and control groups for employment status, where employed learners had 23 percentage point lower likelihood of being in the treatment group (significant at the 5% level). There was also a difference by sex, where being female was associated with an 8-percentage point higher likelihood of being in the treatment group, but this was only marginally significant ($p < 0.100$) and not statistically significant at the 5% level. Both of these observed characteristics were controlled for in the impact estimations. There were no statistically significant differences between the treatment and control groups across age, ethnicity, deprivation or ill-health status.

Additionally, there was no evidence of 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, uncertain, and not statistically significant. This supports the conclusion that the two groups were comparable prior to the intervention. Therefore, even though there were differences in certain observed characteristics, taken together, these results suggest that the randomisation achieved good balance overall. The main analysis was adjusted to control for the characteristics that showed imbalance.

3.2 Outcomes and analysis

3.2.1 Primary outcome analysis

The primary outcome of this trial was whether learners passed or failed the FSQ exam. Table 8 compares the number of learners who passed in each trial arm. Out of 204 learners in the treatment group, 110 passed, giving a success rate of 54%. The success rate was slightly lower in the control group, where 252 out of 515 learners passed, resulting in a rate of 49%. Overall, approximately half of all participants passed across both treatment and control groups. When comparing the groups, the unadjusted pass rate was higher by 5 percentage points in the treatment group than in the control group.

Table 8: Achievement of primary outcome by trial arm

Number of learners	Treatment	Control	Total
Achieved primary outcome	110	252	362
Total number of learners	204	515	719
Percentage of learners who achieved	54%	49%	50%

Note: Source: ILR. Base: All learners as randomised, 719

Table 9 presents the regression results for the primary outcome. 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). This model accounts for the fact that learners within the same providers may share similar characteristics, hence there may be clustering of outcomes at provider level implying that the baseline odds of achieving an outcome may vary by provider. To account for this the model allows each provider to have its own baseline odds of learners passing the exam.

Further, standard errors were clustered at the provider level to ensure that statistical inference remained valid even when observations within providers were correlated. The model also adjusted for region fixed effects since randomisation was stratified by region.

Table 9 presents both unconditional (without controlling for any relevant background characteristics) and conditional (controlling for all relevant background characteristics) estimates. The conditional estimated treatment effect in Model 2 was negative, but not statistically significant (coefficient = -0.05, p = .958), with a 95% confidence interval spanning from -1.94 to 1.84. This wide range indicates a high level of uncertainty around the estimate. The model excluded observations in the North East of England since no learners passed the exam in this region, making within-region comparison impossible. Additionally, age had a positive effect on the outcome, meaning older learners had a higher likelihood of passing (Table A6 - 4 in Appendix 6). With a predicted pass rate of 45% for the control group and 44% for the treated group, this represents a decrease in the probability of passing FSQ Level 1 maths of approximately 1 percentage point. However, this effect is not statistically significant, which means the difference cannot be confidently attributed to the treatment.

Table 9: Regression results from primary outcome analysis

ITT estimates for primary outcome	Coefficient	SE	P-value	CI Lower (95%)	CI Upper (95%)	Base N
Model 1: unconditional estimates	-0.04	0.94	0.967	-1.88	1.80	708
Model 2: conditional estimates	-0.05	0.97	0.958	-1.94	1.84	658

Note: Significance Levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors clustered at provider level and both models have region fixed effects given region is the strata. Source: ILR. Base: All learners as randomised

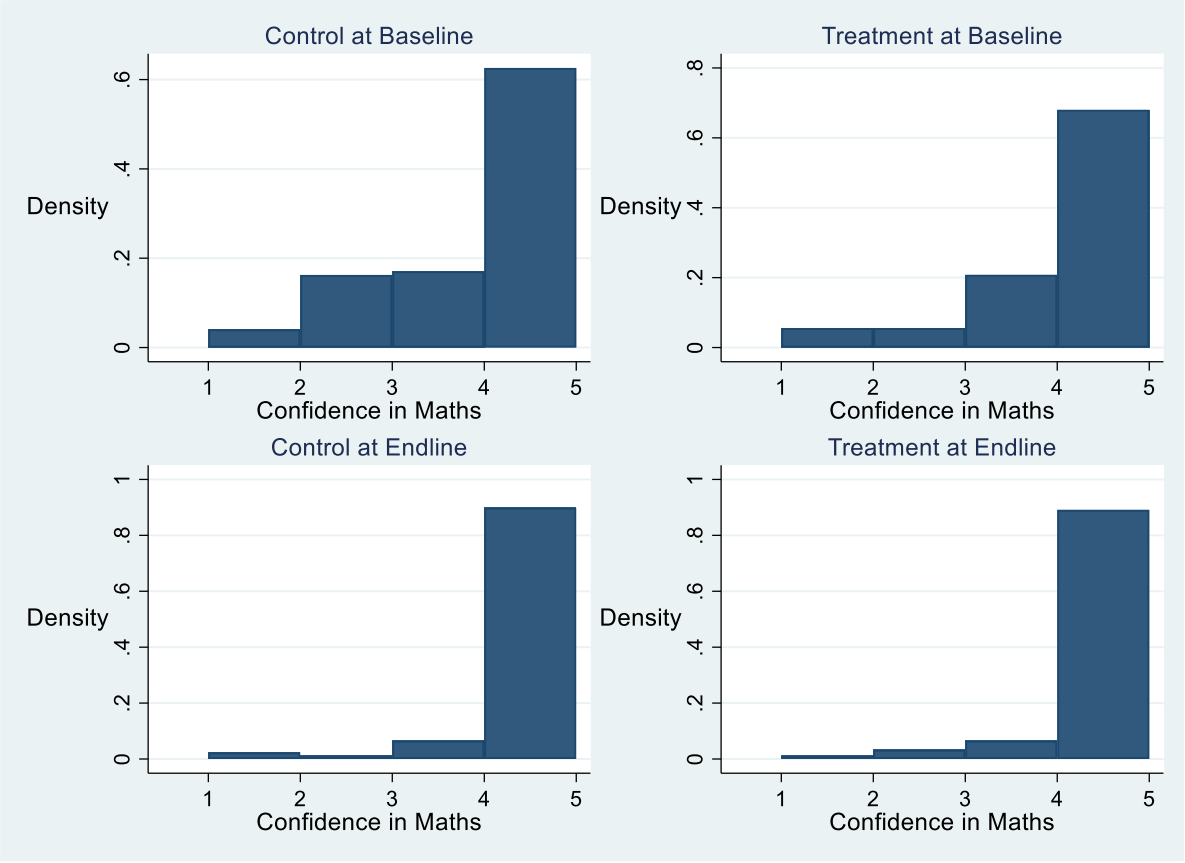
3.2.2 Secondary outcome analysis

The secondary outcome for this trial was learners' confidence in maths. Participating learners were asked to provide information about their overall confidence level in mathematics, as well as their confidence across specific everyday financial situations involving maths or numeracy (e.g., "Understanding interest rates on bank statements"). Each question used a Likert-type scale where 1 meant not confident at all and 5 meant very confident. A composite confidence measure was calculated by averaging learners' confidence scores from these specific financial scenarios. Table A6 - 4 in Appendix 6 compares the secondary confidence outcomes of learners in each trial arm. The response rate was higher at endline than at baseline, and higher in the treatment group at both time points. At endline, the response rate among the treatment group was 45%, whilst the control group was 33%.

Figure 3 shows the distribution of learners' confidence in maths separately for treatment and control groups at baseline and endline. At baseline, both groups show a skewed distribution with most responses concentrated at either "fairly confident" or "very

confident". At endline, both groups show even stronger clustering at the highest level of confidence. Overall, both treatment and control groups reported higher mean confidence at endline compared to baseline, and lower standard deviation at endline, which indicates less variation in responses within groups. The distribution for both groups appears broadly similar.

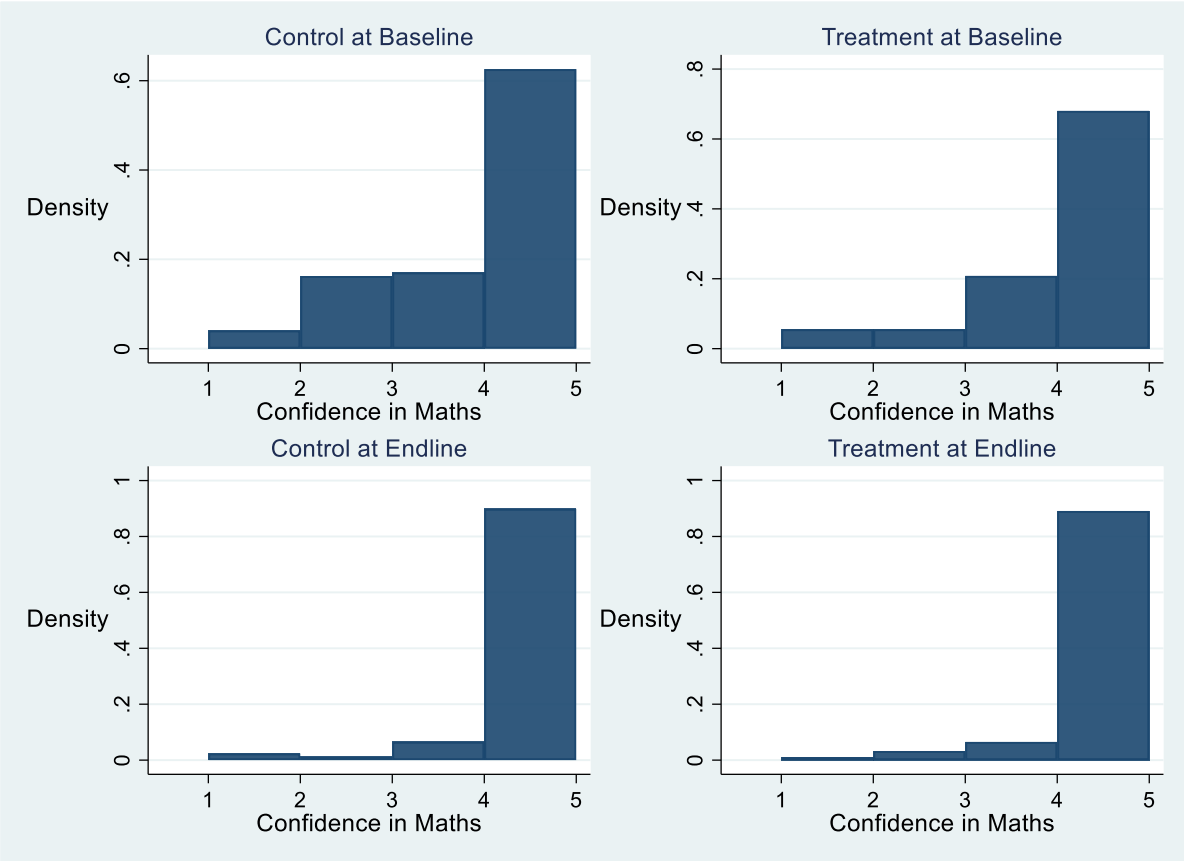
(overall)



Source: Baseline and endline survey. Base: All survey respondents, 195 at baseline and 260 at endline.

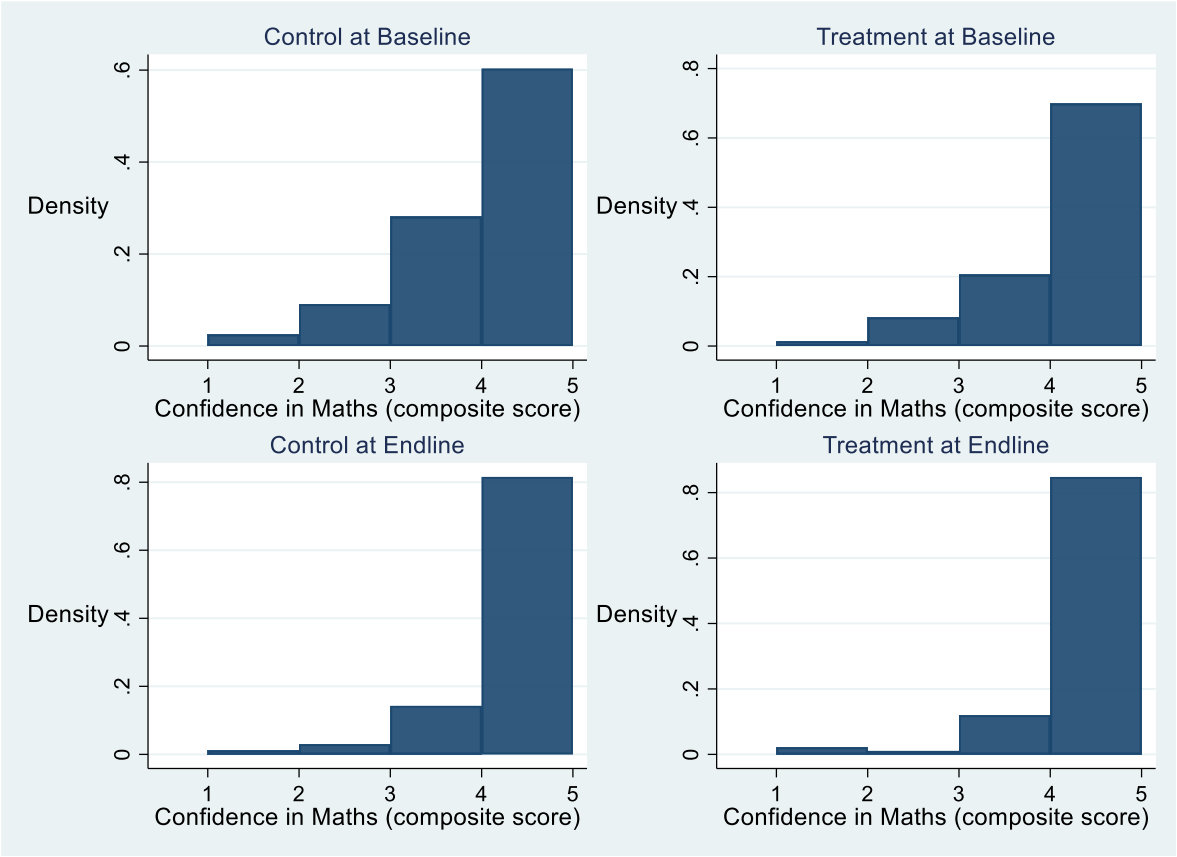
Figure 4: Distribution of confidence in maths (composite score) displays the distribution of learners' mean scores for composite confidence in maths separately for treatment and control groups at baseline and endline. Similar to the first confidence measure, at baseline both groups show a skewed distribution with most responses concentrated between "neither confident nor not confident" and "very confident". At endline, both groups show even stronger clustering at the highest level of confidence. The pattern indicates that average composite confidence increased while the spread of responses narrowed, as indicated by lower standard deviation. The distribution for both groups appeared broadly similar.

Figure 3: Distribution of confidence in maths (overall)



Source: Baseline and endline survey. Base: All survey respondents, 195 at baseline and 260 at endline.

Figure 4: Distribution of confidence in maths (composite score)



Source: Baseline and endline survey. Base: All survey respondents, 194 at baseline and 260 at endline.

Table 10 presents the regression results for the secondary outcomes for both confidence measures. For the overall confidence in maths measure, the estimated effect was negative but not statistically significant at the 5% level (coefficient = -0.17, p=0.223). For the composite confidence measure the estimated effect was positive but also not statistically significant (coefficient = 0.24, p=0.197). These results seem contradictory given that both measured aspects of maths confidence. However, for both measures, baseline confidence was a significant predictor of endline confidence (Table A6 - 4 in Appendix 6). For both results, it is important to emphasise that the low response rate at endline reduced the effective sample size and limited the ability to detect effects reliably and made the results highly uncertain.

Table 10: Regression results for secondary outcomes

Conditional impact estimates for secondary outcomes	Coefficient	SE	P-value	CI Lower (95%)	CI Upper (95%)	Base N
Model 1: Confidence in maths (overall)	-0.17	0.14	0.223	-0.44	0.10	76
Model 2: Confidence in maths (composite)	0.24	0.18	0.197	-0.12	0.60	75

Note: Significance Levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors clustered at provider level, with region fixed effects. Source: Baseline and endline surveys. Base: All respondents.

Furthermore, Table 11 shows results from analysis of each individual question within the composite score which suggests that the positive result for overall composite confidence was likely driven by responses to question one, which was significant at the 5% level (p=0.029) and question five which was significant at the 10% level (p=0.091).

Table 11: Results for each question in composite confidence in maths measure

(How confident do you feel about doing the following things in everyday life?)	Estimate	SE	P-value	CI Lower (95%)	CI Upper (95%)	Base N
Checking your change is right when you have bought something.	0.64*	0.29	0.029	0.07	1.22	72
Working out the best deals when shopping	0.23	0.16	0.152	-0.09	0.56	72
Helping children with maths homework or talking about maths/numbers with children	-0.13	0.15	0.403	-0.42	0.17	38
Understanding interest rates on bank statements	0.26	0.21	0.217	-0.15	0.67	68
Keeping track of your bank account balance	0.25	0.15	0.091	-0.04	0.54	72
Working with numbers as part of a job	-0.41	0.4	0.304	-1.2	0.37	29

Note: Significance Levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors clustered at provider level, with region fixed effects. Source: Baseline and endline surveys. Base: All respondents.

Additional data tables contain more detail on survey responses.

3.3 Additional analyses and robustness checks

Missingness in the primary outcome was overwhelmingly linked to whether learners completed the FSQ Level 1 maths course. Table 12 presents results from robustness

checks for the primary model, which excluded observations that did not have an outcome and were treated as fails. This sensitivity analysis was done progressively, first dropping only the small subgroup of 65 learners who did not have FSQ Level 1 maths as a learning aim in Model 1, then dropping an additional 85 learners who completed the course but still had missing outcomes in Model 2, and finally dropping all observations with missing outcomes in Model 3.

In Model 1, the treatment effect was positive (coefficient = 0.14) as opposed to the ITT analysis (coefficient = -0.05) but was still not statistically significant (p= 0.904), suggesting that excluding this small subgroup did not meaningfully alter the impact estimates. In Model 2, the treatment effect increased further in size (coefficient = 1.50) and became more significant but remained statistically non-significant (p= 0.137) at the 5% level. Model 2 still included 122 learners who did not complete the course or withdrew and had missing outcomes.

In Model 3, which excluded all learners with missing outcome, the estimated treatment effect for learners with outcomes was larger (coefficient =2.05) and statistically significant at the 10% level (p=0.098). However, the confidence interval was wide and included zero, indicating substantial uncertainty around this estimate. Importantly, this effect cannot be directly attributed to the treatment because the learners who had an outcome were also largely learners who completed the course. Therefore, the estimate in Model 3 did not distinguish between the effect of course completion and the treatment itself.

Model 4 uses the CACE/LATE approach, treating random assignment as an instrument. It estimated the causal effect of completing the course among compliers of the FSQ Level 1 maths course. The first stage of this 2-stage analysis provided suggestive evidence that treatment increased the odds of course completion, roughly doubling them on average, which was significant at the 10% level (coefficient = 0.83, equivalent to an odds ratio of 2.3, p=0.071). The second stage results suggested that the treatment effect was not significant among those who complied (coefficient = 0.73, p=0.595). This meant there was no statistical evidence that the intervention affected passing once the effect of course completion was taken into account.

Table 12: Results for primary outcome – robustness check

Outcome	Estimate	SE	CI Lower (95%)	CI Upper (95%)	P-value	N
Model 1: Sample excludes learners who did not have FSQ Level 1 maths in Aims	0.14	1.16	-2.14	2.42	0.904	616
Model 2: Further, sample excludes learners who completed the course but has missing outcome	1.50	1.00	-0.47	3.46	0.137	538
Model 3: Sample excludes all learners with missing outcome	2.05	1.24	-0.38	4.48	0.098	420
Model 4: CACE/LATE stage 1 treatment effect on course completion	0.83	0.46	-0.07	1.73	0.071	623
Model 4: CACE/LATE stage 1 treatment effect on outcome among compliers	0.73	1.38	-1.97	3.44	0.595	616

Note: Significance Levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors clustered at provider level, with region fixed effects. Source: ILR. Base: All learners.

3.4 Missing data analysis

The estimation of the primary outcome and relative covariates used administrative data sources, so missingness was expected to be low. The missingness among the covariates was below the <5% threshold. However, 38% of primary outcomes were recorded as missing. The attrition rate for survey-based secondary outcomes from baseline (considering only learners who responded at baseline) was 58%. Therefore, the analysis explored whether missing data were random or predicted by the treatment assignment or other covariates.

Table A6 - 7 in Appendix 6 reports the results from logistic regression to identify whether missing data occurred at random (MAR) or which (if any) baseline characteristics predicted attrition rates. Missingness in the primary outcome was largely driven by provider differences. A total of 8 providers had no primary outcomes registered for any learner. Additionally, the model omitted 12 providers from the analysis due to lack of variation as they had very few missing outcomes. There were two providers from the

North East who also did not report primary outcomes for any learner and were therefore excluded from the analysis.

Among the remaining 17 providers and 575 learners, ethnicity predicted missingness, with learners from White backgrounds being significantly more likely (p=0.026) to have a missing outcome, with the odds being 1.7 times higher than for all other ethnic groups. Providers who chose the CA Gateway option were significantly less likely (p<0.001) to have missing primary outcomes, with the odds being approximately half.

Other baseline characteristics did not significantly correlate with missingness. Compared to London, learners from the East of England and North West had significantly higher chances of having a missing outcome, both significant at the 1% level (p<0.001). Learners from East Midlands were also significantly more likely to have a missing outcome, significant at the 10% level (p = 0.067). In contrast, learners from the South East and South West were less likely to have a missing outcome, significant at the 1% (p<0.001) and 5% (p=0.039) levels respectively.

Overall, after controlling for covariates, region, and provider differences, treatment was not associated with missing primary outcomes, indicating no differential attrition across treatment groups, whereas ethnicity, provider preference, region, and specific providers showed meaningful associations.

Importantly, missingness in the primary outcome was strongly correlated with FSQ Level 1 maths course completion. Out of 130 learners who did not complete the course or withdrew, 122 had missing outcomes. However, course completion was not a baseline characteristic and therefore not included in the MAR analysis, which restricts predictors of missingness to pre-treatment covariates and stratification factors to preserve the ITT approach. This is explored in the sensitivity analysis in Section 0.

Attrition bias for secondary outcome measures was assessed only for learners who responded to baseline confidence questions in the survey. Results suggest that attrition for secondary outcomes (confidence measures) differed across treatment groups, with assignment to treatment being associated with lower odds of non-response at endline. This may have been due to greater engagement of the treatment group in the trial. In both cases, the association was significant at the 1% level (p<0.001). Learners within providers who selected the CA Gateway option had significant lower non-response rates (p<0.001), indicating better retention when provider preference was aligned.

Learners from the East Midlands and North West of England were significantly less likely to drop out compared to London. This was contrary to the pattern observed for the primary outcome. Regarding baseline data only, missingness levels varied greatly between variables, with higher missingness observed for the confidence in maths survey outcome, where over three-quarters of observations were incomplete. Multiple Imputation (MICE) was used to address missing covariates. Table 13 shows the number of imputations conducted using MICE (Table A6 - 8 in Appendix 6).

Table 13: Missing data imputation – Baseline covariates

Variable	Complete	Incomplete	Imputed	Total
Confidence in Maths	195	524	524	719
Confidence in Maths (composite)	194	525	525	719
Sex	705	14	14	719
Employment	692	27	27	719
Ill-health	698	21	21	719
Ethnicity	691	28	28	719
Deprivation	704	15	15	719

Note: Source: ILR and baseline survey. Base: All learners

The imputation-based outcome analysis is shown in Table 14. The analysis of the primary outcome (attainment) resulted in an estimated treatment effect that was higher than the CACE estimate but not statistically significant (coefficient=0.03, p=0.827). This suggests that results were not influenced by analysing only complete cases, although the imputed estimates remain uncertain. Additionally, imputing the missing covariates did not change the substantive conclusion that the estimated treatment effect on attainment was negligible in magnitude and not significant. The point estimate for the overall confidence measure improved (smaller negative value), and the composite confidence measure also improved once baseline data were imputed. However, neither was statistically significant.

Table 14: MICE - Impact estimate for Contextualised Curriculum

Adjusted impact estimates	Coefficient	SE	P-value	CI Lower (95%)	CI Upper (95%)	Base N
Model 1: Attainment (pass/fail)	0.03	0.13	0.827	-0.23	0.28	719
Model 2: Confidence in Maths	-0.02	0.12	0.835	-0.25	0.20	260
Model 3: Confidence in Maths (composite)	0.07	0.15	0.646	-0.22	0.35	260

Note: Significance Levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Source: ILR and survey.

3.5 Compliance

Compliance among the treatment group was measured by the total number of sessions that each learner attended “fully” or “partially”. It was also assumed that the control group remained unaffected, since randomisation was at the provider level with minimal chances of contamination between treatment and control groups.

Out of 204 learners in the treatment group, only 52% met the minimum dosage requirement of attending at least 10 sessions (80%). Since less than 80% of learners met the compliance threshold, a compliance analysis was conducted. The compliance-adjusted analysis used statistical methods to account for differences in attendance levels. The CACE estimate in Table 15 shows that the effect was positive for learners who completed the minimum attendance (coefficient = 0.01) for the primary outcome, although the p-value was very high and far from being statistically significant (p=0.977).

For both measures of confidence in maths, the direction of the effect remained the same, and overall results remained largely similar both in terms of the size of effects and significance levels. These results suggest that adjusting for compliance does not meaningfully affect the estimated effects of the intervention on either the primary or secondary outcomes. Table A6 - 6Table A6 - 10 in Appendix 6 presents full results.

Table 15: Regression results for CACE

Adjusted impact estimates	Coefficient	SE	P-value	CI Lower (95%)	CI Upper (95%)	Base N
Model 1: Attainment (pass/fail)	0.01	0.25	0.977	-0.48	0.49	669
Model 2: Confidence in maths	-0.22	0.16	0.161	-0.54	0.09	76
Model 3: Confidence in maths (composite)	0.31	0.26	0.231	-0.20	0.83	75

Note: Significance Levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors clustered at provider level, with region fixed effects.

3.6 Subgroup analysis

The analysis examined whether the effect of the treatment differed across sex, ill-health, ethnicity, employment status and learner deprivation by testing for interactions between treatment assignment and each subgroup characteristic for the primary outcome, as shown in Table 16. Across all subgroup analyses, there were no statistically significant effects or evidence that the treatment worked differently for different groups.

The point estimates suggest that the treatment effect might have been larger for females (coefficient =0.49), learners with ill-health (coefficient =0.80), and learners from White backgrounds (coefficient =0.20). They also suggest that the treatment effect might have been smaller for learners who were employed (coefficient = -0.13) and learners living in the most deprived areas (coefficient = -0.26). However, the confidence intervals were wide and included zero for all subgroups, indicating substantial uncertainty and no reliable evidence that the treatment effect differed by these characteristics. Although the

treatment effect did not vary substantially by age, age itself had a positive and statistically significant effect on the probability of passing the course (Table A6 - 5 in Appendix 6).

Table 16: Regression results interaction effects for primary outcome

Interaction effects	Coefficient	SE	P-value	CI Lower (95%)	CI Upper (95%)
Treatment*Female (ref: Male)	0.49	0.62	0.429	-0.73	1.71
Treatment*Age	0.01	0.02	0.666	-0.03	0.04
Treatment*Ill-health (ref: none)	0.80	0.58	0.166	-0.33	1.93
Treatment*White (ref: other ethnicities)	0.20	0.60	0.739	-0.98	1.38
Treatment*Employed (ref: not employed)	-0.13	0.43	0.755	-0.98	0.71
Treatment*Most deprived	-0.26	0.44	0.554	-1.13	0.61

Note: Significance Levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors clustered at provider level, with region fixed effects. Base, n=658

3.7 Effect size estimates

Effect size (Cohen's h) for the primary outcome was calculated using the conditional Model 2 in Table 9, which included relevant covariates. The results in **Error! Reference source not found.** show an effect size of -0.02. The analysis estimated this effect size using a control group probability of 44.6% and a treatment group probability of 43.8%, representing a decrease of 0.8 percentage points in the treatment group.

The effect sizes for confidence in maths were calculated using the square root of the total variance. These effect sizes were substantially larger, with overall confidence in maths showing an effect size of -0.33 standard deviations, and the composite measure showing an effect size of approximately 0.40 standard deviations. The effect sizes from simpler models without covariates were very similar but are not reported here.

Table 17: Results – conditional effect sizes

Outcome	Effect size	ES CI Lower (95%)	ES CI Upper (95%)
Attainment (Cohen's h)	-0.02	-0.12	0.09
Confidence in Maths	-0.33	-0.87	0.20
Confidence in Maths (composite)	0.40	-0.21	1.00

Note: Source: ILR and survey. Base: All learners

3.8 ICC Estimation

Error! Reference source not found. reports the Intra-class Correlation Coefficient (ICC) estimation results. The primary outcome conditional Model 2 in Table 9 was used to calculate the ICC. Results suggest that about 42% of the total variation in the primary outcome results is explained by provider-level clustering. This meant the training provider a learner attended had a substantial impact on their outcome and indicated considerable similarity in outcomes for learners within the same provider. Taking into consideration the high ICC of 42%, an average cluster size of 25, an assumed target effect size (Cohen's h) of 0.2 and the 42% provider-level attrition observed in the pilot trial, the ex-post power calculation (Appendix 6) showed that approximately 600 providers would be required to run a full-scale trial to detect a meaningful effect. However, given the near zero effect size observed in the pilot trial and the pooled standard deviation of 0.5, the number of providers required would be impractically high (in the thousands), making a full trial unfeasible.

Table 18: ICC estimation results

Outcome	Cluster Variance	Residual Variance	ICC
Primary Outcome	2.39	3.29	0.42

The trial had 29 clusters (providers) post-attrition and an average cluster size of 25. With an ICC of 0.42, at 80% power the MDES[13] (Cohen's h) it would have been able to detect is approximately 0.7.

[13] 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.

4 Implementation and process evaluation results

This chapter draws on qualitative and quantitative data from interviews, surveys, and attendance records to examine how CA was delivered by tutors and experienced by learners. The evaluation considers tutor and learner engagement, delivery practices, and contextual factors influencing participation. The analysis highlights key aspects of delivering trials, covering compliance, attrition, fidelity, feasibility, and readiness for future trial activity. It offers an account of delivery and practical considerations for future implementation and scale-up. Key IPE findings are provided below, with further detail provided throughout the chapter.

- **Compliance**: Compliance with the required data uploads varied considerably. The administrative demands of the trial were off-putting for some providers who withdrew. 10 of the 13 tutors who went onto deliver the CA achieved the threshold for compliance set by the product developer. This meant they attended all training and 80% of the ongoing CPD sessions. More than half (59%) of learners met the minimum requirement of attending at least 10 sessions. This was a similar level of learner compliance as seen in the AMA trial. There was no evidence of contamination between treatment and control allocations.
- **Attrition**: Relatively high numbers of providers dropped out post-randomisation and pre-training.
- **Fidelity**: Tutors who delivered the intervention did so with only minor adaptations to better support learner needs, mainly adjusting the pacing the content. Taking part in ongoing CPD sessions helped them deliver the intervention in line with expectations, following the structure and building learning through interactive sessions, peer learning and problem-solving.
- **Feasibility**: Overall, the findings suggest the intervention is feasible to deliver. Whilst some barriers and potential improvements were identified, tutors addressed these through appropriate adaptations.
- **Readiness for trial**: Based on IPE evidence, the intervention appears ready for wider rollout. Additional guidance could be provided on group size, curriculum alignment, and support for ESOL learners. Training formats could be more flexible to accommodate part-time staff. Tutors expressed a preference for more comprehensive FSQ curriculum coverage and more structured support through planned check-ins and observations of practice.
- **Perceived outcomes**: Tutors reported positive impacts on their professional development, noting that the approach prompted them to reflect on their practices and make lasting changes to their FSQ Level 1 teaching. Learners in the treatment group consistently reported in interviews that embedding maths in real-life contexts enhanced their understanding of the relevance and practicality of maths.

4.1 Compliance

Of the 13 tutors who delivered the CA course, 10 met the compliance threshold by completing the full pre-training and attended at least 10 study sessions. One tutor completed the full pre-training but only attended 9 study sessions. The remaining 2 tutors received partial or no training but still attended 10 and 9 study sessions respectively (Table 19).

Tutors with lower attendance at training or lesson study sessions cited scheduling and communication challenges**.** Part-time tutors faced particular difficulties due to late notification of training sessions and timings that fell outside their contracted hours. In some cases, managers attended on their behalf and shared notes. Providers also noted that late notification of training, and notification during the busy exam period in July was not ideal. Those recruited late to the trial received information about training at the end of July or early September. Additionally, trial readiness packs arrived in September, very shortly before the planned start. These factors complicated organisation, planning and cover arrangements. Some tutors who were unable to attend training sessions would have liked access to recordings, particularly those who were part-time:

> With the initial information, we signed up, and there was a silence for quite a chunk of time, and then all of a sudden it was like, 'Right, the training sessions are here, and you need to do them next week.' That doesn't work for us. It was exam time, all of our staff were busy, and they were already allocated into slots. And like I say, we only have a few maths tutors here, we're a very small team, and none of them are full-time. And we don't have staff available to cover, really. So, it just wasn't feasible for us to get staff onto those sessions. - *Tutor, Treatment group*

Table 19: Tutor training Status

Tutors Status	Number of tutors
Fully trained and complete study sessions	10
Fully trained and partial study sessions	1
Partially trained and complete study sessions	1
No training and partial study sessions	1
Total	13

Note: Source: programme management information

The delivery of the CA FSQ sessions varied across tutors, with some not delivering any sessions as planned. Table 20 shows the number of tutors who delivered each session. A total of 206 sessions were planned. However, the evaluation did not have data for 2

sessions, so it is unclear whether these sessions were delivered at all. Out of the remaining 204 sessions, all 13 tutors delivered each session for at least 1 cohort, and 4 tutors delivered each session for an additional cohort. However, 3 out of these additional sessions were not delivered, despite being planned, resulting in 201 out of 206 sessions being delivered (98%).

Table 20: Number of tutors who delivered each session

Sessions	No. of tutors who delivered the session in at least 1 cohort	No. of tutors who delivered a second cohort of same session	Total
Session 1: Percentages 1	13	4	17
Session 2: Percentages 2	13	4	17
Session 3: Percentages 3	13	4	17
Session 4: Ratio tables 1	13	4	17
Session 5: Ratio tables 2	13	4	17
Session 6: Data 1	13	3	16
Session 7: Data 2	13	3	16
Session 8: Geometry 1	13	4	17
Session 9: Geometry 2	13	4	17
Session 10: Geometry 3	13	4	17
Session 11: Fractions	13	3	16
Session 12: Ratio	13	4	17
Total	156	45	201

Source: Data submitted to the Ipsos Data Portal by providers

Table 22Table 21 shows learners' attendance by session. Across the first 6 sessions, the total reported number of learners are 196, meaning eight learners did not attend any sessions at all, these further drop to 189 in the last 6 sessions. Attendance was highest in the early sessions, with 74-84% of learners attending full sessions. Attendance declined over time. For sessions 7-12, 65-74% of learners attended full sessions and the highest non-attendees in session 10.

Table 21: Learner attendance by session

Session	Attended all (N)	Attended all (%)	Attended part (N)	Attended part (%)	Did not attend (N)	Did not attend (%)	Total
Session1	150	77%	6	[u]	40	20%	196
Session2	149	76%	11	[u]	36	18%	196
Session3	149	76%	10	[u]	37	19%	196
Session4	164	84%	9	[u]	23	[u]	196
Session5	151	77%	9	[u]	36	18%	196
Session6	145	74%	6	[u]	45	23%	196
Session7	123	65%	12	[u]	54	29%	189
Session8	136	72%	8	[u]	45	24%	189
Session9	139	74%	6	[u]	44	23%	189
Session10	123	65%	10	[u]	56	30%	189
Session11	136	72%	12	[u]	41	22%	189
Session12	147	78%	10	[u]	32	17%	189

Source: IPSOS learner data. Base: All learners in treatment group, 204
'Attended part' means that a learner attended part of that session. Shorthand [u] is used in place of %s where cell counts are low (<30)

Table 22 shows learners' attendance at sessions. Out of the 204 learners in the treatment group 196 attended at least one session. Approximately 59% of learners met the minimum threshold (the dosage requirement) set by the product developer, of attending at least 10 sessions (80%). This was similar to the level of attendance in the AMA trial. A key requirement of the trial was that participating learners had to be aged 19 by 1 September 2024 to be eligible. However, some providers enrolled learners who were under the age of 19. To ensure compliance with the trial requirements and data sharing arrangements, these learners – a relatively small number – were removed from the data.

Table 22: Learner attendance

Session attendance	Frequency	Percentage
Did not attend any session	8	[u]
Did not attend minimum dosage	76	37%
Attended minimum dosage	120	59%
Total	204	100%

Note: shorthand [u] has been used to indicate well percentages have been suppressed due to low reliability i.e. cell counts <50

Source: Learner data submitted to the Ipsos Data Portal by providers

Attendance levels were affected by personal barriers, rather than low engagement with the intervention. Interviews from tutors and learners indicated that learners were generally positive about the CA. The interviews also highlighted a range of personal barriers that influenced lesson attendance, such as work and family demands which are common for adult learners. Examples included health issues, children being unwell, holidays and other prior commitments, job interviews, and bereavement, suggesting that lessons were missed mainly due to personal circumstances rather than lack of engagement with the intervention itself.

All participating providers were required to upload data on learners taking part in the trial to the Ipsos Data Portal. This included learner demographic data, contact details and, for those in the treatment group, attendance at each of the 12 CA sessions. Providers were also required to confirm that learners had been provided with information about the trial and given the opportunity to opt-out of participation. Providers were asked to submit their learner data within 2 weeks of the start of the course, although most submitted later than this deadline. Nevertheless, all providers participating in the trial had successfully submitted their learner data by July 2025. Learner data uploads were typically completed by a strategic lead within the provider, such as a curriculum manager, sometimes with support from a data manager.

Most providers submitted data to a high standard of quality and completeness, despite initial inconsistencies and delays. This was due to the inclusion of compulsory fields and validation rules within the data submission template, which were designed to ensure quality and accuracy. However, it did result in some providers submitting default entries for some mandatory fields, in particular date of birth, suggesting that additional checks should be completed should this approach be used in future. The only missing data were telephone numbers, which was missing for 1% of learners, and email addresses (missing for 2%) as these were not compulsory fields.

Some providers faced challenges in uploading their learner data. These challenges included lack of experience with data and spreadsheets, error messages when attempting to upload due to data being missing or in the wrong format and confusion about the requirements. These issues were resolved through dedicated support provided by Ipsos via email and 1-to-1 calls with providers.

4.1.1 Tutors' views on the intervention training

Some tutors valued the training for its clarity, interactivity, and relaxed delivery. They described it as informative and well-delivered, highlighting the interactive format and contextualised strategies. The relaxed training style supported their understanding of how the strategies could be applied in real classroom settings by allowing a flexible pace, which gave tutors time to reflect on and explore strategies without feeling rushed.

> I leave the sessions having a good understanding of what we have set out to accomplish. It is interactive and relaxed, but, at the same time, informative. - *Tutor, Treatment group*

Tutors deepened their understanding of mathematical thinking by reviewing maths questions from a learner's perspective. The design of the training sessions helped them anticipate diverse strategies and starting points when facilitating learner problem-solving discussions.

> We did the questions ourselves, but we also thought about how our learners might approach it... They may not know things that I know, it was useful to do that - *Tutor, Treatment group*

In interviews, tutors described the training as high-quality and well-paced. They noted that it built their confidence to deliver the intervention. Most felt well-prepared to deliver the CA, with core principles having been clearly communicated.

> I think straight away after having training, I felt quite confident - *Tutor, Treatment group*

Initial concerns about the workload eased as some tutors learned new strategies and saw benefits for their teaching. Materials were described as thorough and manageable within existing planning time. Some tutors who joined the July training felt better prepared than those who joined the September training, as they had more time over the summer break to plan how the approach would fit with their level 1 FSQ teaching, and to organise resources and lesson materials. The varying start starts across providers created uncertainty, with some tutors feeling behind as they had yet to start teaching sessions whilst others had already begun delivery. They felt unable to make the most of the lesson study sessions and requested clearer guidance on when to begin delivery. While this occasionally limited the depth of the lesson study discussions, the sessions remained useful for anticipating learner responses and understanding teaching approaches.

Tutors in interviews frequently said they found the lesson materials to be clear, and they felt well-equipped to deliver the approach in the classroom. They consistently praised the completeness of training resources, including lesson plans, slides with accompanying notes, despite some finding it hard to fit everything into their courses in full. The clear breakdown of materials and resources for each of the 12 lessons was seen as particularly helpful for preparation. The slide notes were especially helpful in preparing lessons and facilitating learner-focused problem-solving.

> We've been given this lovely material…the PowerPoints are all there, we've got all the notes…it's almost like a dot-to-dot lesson. - *Tutor, Treatment group*

4.2 Attrition

Management information data shows that 42% of providers (21 out of 50) and 50% of tutors (20 out of 40) withdrew before delivery began or in the early stages of delivery. In the time between expressing interest and being randomised to delivery, there was relatively high attrition among the providers. This pattern may suggest challenges with clearly communicating the delivery details.

However, interviews with tutors and providers who remained in the trial revealed some logistical and communication issues which may have contributed to trial attrition. These issues centred on the volume of communications and late delivery of trial readiness packs. For example, a provider manager in an organisation that signed up for several Multiply trials reported they had been unaware of the need for tutors to attend lesson study sessions due to the high volume of information they received. In other cases, staff had not engaged with the trial readiness materials, either because these were not shared by senior managers, were sent to a different contact in the setting and not passed on or had simply not been received. These communication gaps may have contributed to gaps in understanding and delivery preparedness.

However, after this point attrition within the treatment group was low, as was the case with the control providers. Providers in the control group were generally content to be part of the control group and saw the value of participating in the trial. While some control providers found the administrative burden heavy, the funding helped to cover this. Some control group providers expressed concern that had they been in the treatment group, it would have been difficult to deliver due to the short timeline between sign-up and the intervention start date.

Providers who withdrew from the trial, whether before randomisation or after, did not always give a reason in the information held by the MSS. Reasons given included lack of buy-in from tutors, lack of staff capacity for training or delivery, or being unable to recruit enough learners.

4.3 Fidelity

This section describes the extent to which the intervention was delivered as intended by treatment providers. It explores some of the key assumptions identified in the Theory of Change, providing evidence on the training and support delivered to and taken up by tutors, and examining the teaching delivered to learners and the extent to which this teaching reflected the intended approach.

Interviewed tutors commonly adapted the CA to meet learner needs. Whilst this sometimes led to partial implementation of the intended model, overall, the evidence suggested tutors maintained a good degree of fidelity. Tutors demonstrated an understanding of the core principles of the CA and made efforts to incorporate them into their teaching. Interview data suggested that delivery was often adjusted in response to

learner needs and the time constraints of the lessons. Specifically, most tutors delivered all 12 lessons, but some adjusted the pacing and content, particularly for larger or mixed-ability groups, to ensure lessons could be completed within the 90-minute timeframe. Learners needed time to engage with the approach. It was common for tutors in interviews to report needing to balance delivery with exam-related content:

> Sometimes sticking to an hour-and-a-half and not going over can be quite difficult… you have to do the whole of reigning that back in again to carry on - *Tutor, Treatment group*

Some tutors adapted delivery by carrying over content between sessions and streamlining activities, including skipping some examples. Though time-intensive, the CA was perceived by tutors and providers to foster deep learning, allowing learners to work at their own pace. Teaching experience, group size, and learner needs shaped how tutors managed timing across lessons. Tutors responded thoughtfully to these dynamics, adapting their approach to keep learning inclusive and responsive to learner needs.

Some tutors reported that variable learner attendance affected learning and disrupted the continuity of lessons. Learner attendance was relatively high with 59% completing the number of sessions judged necessary to deliver the intended outcomes from the intervention (Table 22). Some said it was challenging to support catch up for learners who had missed sessions while at the same time facilitating new content and group work. This could lead to lessons being rushed. In one example, this was necessitated by a rolling enrolment approach, which meant new learners had to be accommodated mid-course. The intervention was not designed to accommodate this type of course delivery. Despite these challenges, tutors made efforts to bring learners up to speed through 1-to-1s or by assigning them extra coursework.

4.3.1 Learner reflections on delivery

During interviews, some learners recalled elements of the CA, though not all elements, possibly because they were integrated into the learning experience rather than explicitly identified as part of the approach. Reflecting on their learning experiences, they highlighted the use of real-life examples, physical objects and visual representations to help them engage with abstract concepts through real-world contexts, as well as group work and peer learning, exploring multiple problem-solving strategies, and interactive board work. Compared to control group learners, treatment group learners mentioned a wider range of collaborative activities:

> We get time to interact with each other in the groups to bring out answers, showing our whiteboards. It's a really good move, really interactive. – *Learner, treatment group*

Some learners struggled to balance the volume of homework with work and family life. This was exacerbated when tutors assigned extra homework to cover content missed in

lessons or exam-style questions with FSQ tests outside the CA curriculum. Despite these challenges, most learners found the homework beneficial.

> They were supposed to be 1-hour long sessions but normally, it will take me 2 hours, and we were not able to complete the work sheets in the class. I would give it as homework – *Tutor, treatment group*

4.3.2 Comparison with business as usual (BAU)

The CA was unfamiliar to most tutors who delivered it and represented a shift from their usual teaching practice.

Before implementing the intervention, many treatment group tutors reported in interviews that they already used real-life examples in their teaching, which overlapped with some aspects of the CA. However, they described a more superficial use of realistic contexts, aimed at making maths relatable, rather than supporting deeper engagement with mathematical principles. They acknowledged that the intervention went further, placing emphasis on learner-led exploration and reasoning before introducing formal methods. This distinction was also evident among some control group tutors, who said they were familiar with contextualised teaching but primarily associated it with using real-life scenarios. Their descriptions indicated that they did not use the CA approach:

> So, I know a bit about the importance of contextualising maths. I'm not sure if I can talk about a particular approach that you might be referring to. I've come across it on various training courses. – *Tutor, control group*

Whilst learners and tutors in the control group referred in interviews to the use of real-life examples, practical contexts and group work, these were not consistent features of their courses. There was little evidence of the structured, contextualised principles that characterise CA, such as building on what learners already knew rather than jumping into formulas, encouraging learner-led discussion, or starting from a real-world context.

4.3.3 Contamination across arms

There was no indication that the CA had been unintentionally used in control group classes. Many providers offered only one FSQ course, and so therefore did not have treatment and control group classes running in parallel, which lowered contamination risks. At larger providers, some tutors in the treatment group were sharing resources and ideas informally across the FSQ teaching team, but most were waiting until the trial ended to do this. When sharing of ideas did happen, the interviews suggested it was not detailed enough for others to adopt the approach fully. Nothing from the control group interviews suggested otherwise. Some learners in the treatment group may have experienced aspects of the traditional FSQ curriculum approach, especially since the CA did not cover the full curriculum. However, all treatment group tutors interviewed reported using aspects of the CA throughout their delivery.

4.4 Feasibility

This section explores tutor and learner experiences of the CA, including the enablers and barriers they encountered.

Learner engagement was shaped by individual circumstances, but adaptive teaching strategies could overcome these barriers. Tutors reported that some neurodivergent learners required additional time to fully engage with the classes and adjust to the approach. A tutor who experienced resistance from more confident learners who wanted a set method of learning was concerned that this affected the group's buy-in to the CA. However, tutors were able to tailor their delivery of the CA to support engagement and support the diversity of learner needs. For example, interviewed tutors described how learners with autism found the CA unpredictable and preferred clear, structured instructions. Some tutors also felt that anxious learners were initially reluctant to engage with group work, although over time the approach helped to build their confidence. Some tutors adapted strategies, such as using mini white boards instead of whole-class presentations, to reduce anxiety. They highlighted the importance of differentiation, visual aids, and step-by-step instructions to support these learners effectively.

English for Speakers of another Language (ESOL) learners experienced language barriers but were supported through visual aids. ESOL learners from both control and treatment groups reported difficulties with language, but treatment tutors said this was more of a challenge with the CA, which requires more reading. Visual aids were considered an effective enabler for this group. Some tutors also noted the need for simpler language, shorter examples, visual aids, and language scaffolding such as key vocabulary and sentence starters.

Some tutors and learners reported that group work and peer support enabled engagement and supported those with different learning preferences or ESOL needs. For example, some tutors described how pairing ESOL and SEND learners with supportive peers helped overcome reading and comprehension barriers. Some learners with different abilities valued peer support and gave examples of helping each other, with some citing the presence of classmates as a motivator for attendance.

> If I didn't understand it and they did, you know, I'll ask them, 'Did you get this bit, or did you get that bit, or what were you stuck on there?' And then we, kind of, work together, that was nice. It's that support unit that we had. - *Learner, treatment group*

Some tutors made minor adaptations to delivery to meet the diverse needs of learners, particularly those with language barriers, limited prior education and those returning to education. Strategies included simplifying language, using more visual and hands-on tools, and introducing paired discussion to support participation. Some tutors observed that adults returning to education after a long gap were more open to trying new techniques than other learners.

> Adults, I think they are more open to different concepts. And also, when they were at school, let's say, 15, 20 years ago, again, things were different as it is now. And as I've said, they've forgotten all of it, so then for them, it's all new anyway - *Tutor, Treatment group*

In contrast, some tutors said that younger or more confident learners, who were accustomed to more traditional maths teaching, preferred direct instruction. The CA was perceived as an obstacle for these learners, who saw its methods as inefficient and slow. Over time, some of those learners who were initially resistant warmed to the intervention once they understood its benefits.

Some tutors adapted delivery to improve accessibility for ESOL learners. They simplified language, avoided or explained culturally specific references, incorporated paired discussion and vocabulary scaffolds, and used visual or hands-on strategies to support understanding and encourage participation.

4.4.1 CPD Lesson study and peer collaboration

The collaborative lesson study sessions provided to tutors during delivery of the intervention strengthened their facilitation strategies and built their confidence to deliver with fidelity through shared reflection opportunities. Some tutors consistently valued the opportunity to exchange strategies, challenges, and delivery experiences through whole-group peer discussion. These reflective spaces helped them refine their practice and adapt strategies to real classroom contexts. The repeated cycle of planning, delivery and review supported deeper engagement with the approach and greater consistency in delivery. This was despite some not understanding the number and duration of these ongoing support sessions.

Techniques such as ‘thinking time’ and ‘pass the pen’ supported learners' reasoning and problem-solving, according to some tutors. The shift from directive teaching to learner-led facilitation was unfamiliar and occasionally challenging, as this was a new approach to teaching for many. Tutors reflected on balancing discovery-based learning with timely input to manage their classes, which was something the intervention training had provided limited guidance on. The CA was demanding in terms of group management, requiring tutors to enable all voices to be heard whilst keeping learning focused.

> A big part of the lessons is meant to be that you let the students learn through discovery and you're not meant to input-match. So, there's quite a lot of discussion [at times]. Well, I had to input at that point, otherwise they'd not have gotten any of it, or I had to stop that person then. - *Tutor, Treatment group.*

Trainers observed some tutors gaining confidence from the lesson study through peer discussions, which often sparked ‘lightbulb moments’.

> You could see them kind of having a few lightbulb moments, where they were like my gosh, lots of people did say that. - *Trainer, Treatment group*

The trainers observed that some tutors were excited by the CA, bringing learners' work into lesson study sessions for discussion. Whilst some tutors expressed initial resistance to the CA, most became more engaged as they saw its impact on learner confidence and motivation.

Tutors described the lesson study sessions as a source of professional connection. They appreciated the opportunity to connect with peers facing similar delivery challenges, especially those working in smaller teams or across dispersed sites. These discussions helped reduce isolation and build confidence in implementing the CA strategies.

Lesson study sessions supported delivery and addressed most questions about the intervention. Outside of these sessions, some tutors sought support from line managers largely due to their immediate availability and familiarity. Whilst the trainers remained accessible, some tutors reflected that more structured follow-ups would have been helpful, especially during the later stages of delivery. They suggested that light-touch check-ins might enhance continuity and reduce uncertainty. Trainers also proposed visiting providers to observe classroom delivery and offer targeted feedback and support, which could strengthen the link between training and practice.

Some providers and tutors endorsed the CA, saying they would recommend it to colleagues or other training providers. The ‘ready to go’ lesson materials and tools, such as the ratio table and percentage bar, were valued for their clarity and adaptability to other mathematics courses (pre-entry, level 2, etc). This demonstrated the theory of change a mechanism for contextualisation as a leading device for understanding a topic spreading within institutions.

Tutors offered constructive suggestions to improve future training including:

- Opportunities for live observation or videos showing tutors delivering the teaching strategies in real classroom contexts
- Recording of training sessions to enable part-time colleagues to catch up
- Strategies to support learners who joined courses late, and how best to revisit earlier taught content
- Greater flexibility in training delivery and lesson study through multiple dates and timeslots, especially lesson study sessions during the working day.
- Condensed training formats to ease the start-of-term scheduling pressure, though this might negatively impact training pace.

While some tutors expressed a preference for face-to-face sessions, they acknowledged practical issues associated with the geographic spread of participants. They also highlighted the value of clearer communication regarding session dates, cancellations, and lesson start timelines. Earlier access to materials and guidance on when to begin teaching would help ensure consistency across providers.

4.4.2 Learner experiences and engagement

Among learners who responded to the questions related to satisfaction in the endline survey, there were generally high levels of overall satisfaction with their FSQ Level 1 course (Figure 5). Overall, 93% of learners reported being either 'very satisfied' or 'fairly satisfied' with the course. In response to a series of questions regarding satisfaction with various aspects of the course – such as satisfaction with the content of learning materials, workbooks or reading materials, including length of the each session, length of course, the time of day of lessons, quality of teaching, information provided about other courses, difficulty of maths and satisfaction with overall learning experience – satisfaction levels were generally high among all learners.

Figure 5: Overall, how satisfied or dissatisfied were you with the courses?

Note: Results from the endline survey. Non-responses are excluded. There were 364 missing responses in the control group and 115 in the treatment group.

Learners in the survey generally found the level of maths covered in their course appropriate as opposed to being either too easy or too difficult (Table 23).

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

Level of maths	Control N	Control %	Treatment N	Treatment %
Too easy	21	[u]	12	[u]
About right	118	78%	67	75%
Too difficult	[c]	[u]	[c]	[u]
Don't know/ can't remember	[c]	[u]	[c]	[u]
Total	151	100%	89	100%

Note: Results from the endline survey. Non-responses are excluded. There were 364 missing responses in the control group and 115 in the treatment group. [c] represents count less than 5 that have been suppressed for confidentiality or indicates secondary suppression and [u] indicates percentages with low reliability due to cell counts <30.

Learners responding to the survey generally found their course very engaging (Table 24).

Table 24: Overall, how engaging did you find the course?

Engagement level	Control N	Control %	Treatment N	Treatment %
Very/ fairly engaging	131	88%	81	91%
Neutral	15	[u]	[c]	[u]
Not very engaging/not engaging at all	[c]	[u]	[c]	[u]
Don't know	[c]	[u]	[c]	[u]
Total	149	100%	89	100%

Note: Results from the endline survey. Non-responses are excluded. There were 366 missing responses in the control group and 115 in the treatment group. [c] represents count less than 5 and [u] indicates percentages suppressed due to low reliability, where cell counts <30.

Learners interviewed were generally positive about their CA courses. They particularly appreciated the visual elements, group work, and exposure to a variety of possible methods. Some learners found visual elements, such as percentage bars, engaging and well-suited to their learning styles. Others found demonstrations particularly engaging.

Group work and peer learning built confidence, enabling some learners to hear alternative explanations for solving mathematical problems. Some felt more confident engaging in discussion and asking questions within their peer groups. When multiple methods were used to reach a solution, some tutors would explain each one, allowing learners to choose the one that suited them best, which learners found particularly helpful.

> I think it's really useful, because, obviously, sometimes, one method is not going to work…your brain's not computing the information. - *Learner, Treatment group*

Some learners found the real-life contexts familiar and helpful. They explained that the real-world application enhanced their engagement and made the course feel practical. Examples were often drawn from daily life, including shopping, home improvements and finances. They felt that using real-life examples made concepts easier to understand and memorable.

During interviews, some learners attributed their positive experience to their tutor but did not highlight specific aspects of the teaching approach. Some tutors were praised for their patience, breaking down questions, and giving individual support to those who were struggling. Other learners felt that their positive experience was due to returning to education with an open mindset and greater maturity than they had previously.

However, some learners disliked elements of the CA, particularly those who had enjoyed traditional maths education. Some found the use of multiple methods confusing, especially learners with ADHD. Others disliked learner-led elements of the CA, such as

being asked to write on the board. Those with recent experience of, or those who liked, more traditional maths education was more resistant, finding it difficult to adapt to aspects of the CA.

> A learner with maths qualifications from their home country preferred the traditional approach to maths teaching. They found the demonstration of multiple methods challenging because they preferred to use the method they were already familiar with. They did not know how to approach maths problems using a contextualised learner-driven approach and engaged less with group discussion. *Learner*, *Treatment Group*

Learners reported in interviews that teaching sometimes felt rushed and/or they had too much extra learning to do outside of class. They explained that rushed classes meant they missed some content. This required them to catch up in additional sessions or do extra independent learning in their own time, which was challenging to fit into daily life. A learner who did not pass their exam attributed this to missed content in class.

> She [the tutor] just said she was following the programme, and then, when the programme had finished, there was loads of stuff that we hadn't covered, so she just tried to cover as much as she could, before we had our exam. - *Learner, Treatment group*

Learners suggested improvements that would allow for more content to be covered and to better suit their learning styles. These included extending the course duration by several weeks or increasing weekly hours, allowing more time to cover content. This was especially, but not limited to, ESOL learners. Learners also said that having more frequent classes would help them retain content between classes. Others wanted additional resources for self-study, such as textbooks.

4.5 Readiness for trial

This was the first time the CA had been trialled with adult learners undertaking the Functional Skills Maths Level 1 qualification. Consistent with studies of the approach in primary to Key Stage 1 settings, the qualitative data indicated high levels of engagement and broadly positive perceptions of the approach. There is evidence to suggest that the CA can enhance learner experience, although this varied depending on factors such as tutors' teaching style and their connection with learners. For example, learners often attributed their positive experience to their tutor rather than to specific methods, highlighting the importance of tutor-learner rapport and the need to better distinguish the impact of the pedagogical approach. Future studies could include classroom observations to deepen understanding of this.

The delivery of the CA was broadly feasible in most provider settings. Training materials were clear and easy to implement, and many tutors reported a shift in how they used real-life examples, moving from directive teaching to learner-led facilitation. Whilst some

found this shift challenging, their initial apprehensions reduced as they observed improvements in learners' confidence, motivation and ownership of problem-solving. Tutors valued the professional development opportunity and described the training as well-paced, interactive and useful.

However, the feasibility of delivery was constrained by several factors. Attendance issues, common in adult learning, posed challenges for the full implementation of the intervention. Missed lessons disrupted continuity, with tutors trying to catch learners up whilst deliver new content in the limited timeframe of the lesson, which sometimes made learners feel they were rushed. These issues may have affected tutors' ability to implement aspects of the approach such as facilitated problem solving and group discussions, which require more time than traditional maths delivery. Class size also impacted feasibility, with larger groups needing more active management, making it difficult to sustain discovery-led learning, problem solving and discussion. Smaller group sizes were more conducive to this style of collaborative learning.

Other feasibility considerations included the need for more flexible training formats, earlier access to materials and clearer guidance on delivery timelines. Tutors also highlighted the need to align the intervention with the full FSQ curriculum and provide additional supporting materials for ESOL learners.

Based on IPE evidence, the CA appears to demonstrate potential for wider rollout with evidence of transferability among FE providers and sustained impact on teaching practice. Scale-up would require consideration of staffing ratios, more advance notification of training dates so that these could be built into tutors' schedules, a review and potential redesign of training and CPD to reduce intensity and burden on providers and tutors, and ongoing support through regular check-ins and observations to maintain intervention fidelity.

4.6 Perceived outcomes

4.6.1 Tutors' professional growth and future integration

Interviewed tutors commonly reported that they had gained pedagogical confidence, renewed professional purpose and a flexible toolkit for engaging diverse learners. Many tutors described the CA as refreshing, re-energising their teaching practice and offering tools to re-engage learners. The approach provided practical strategies for engagement through learner-led discussion and problem-solving. A tutor reflected:

> I feel like I'm learning and I'm developing something as well, not just teaching the same thing. - *Tutor, Treatment group*

The intervention was perceived by interviewed tutors as credible and impactful, resulting in sustained changes in teaching that persisted beyond the trial period.

The renewed confidence among tutors led to continued use of CA content beyond the trial period. They selectively used strategies such as ratio tables and facilitated questioning in other maths classes, including GCSE and FSQ Level 2, applying these to

topics like percentages, fractions, and ratio tables. Tutors described adapting aspects of the approach to suit learner needs, blending it with their existing practice and planning to embed it into their long-term teaching.

> I am really enjoying it because it's going to benefit my students. Already I'm thinking, like, 'Oh, how can I just bring some of this into entry level?'…I had a drop-in on Tuesday, and I had 3 level 1s, 1 of my level 1s, and I had a couple of level 2s in there. We were doing the ratio table of doing that, and they're, like, 'Oh my God, we've got this.' This is where I'm going to bring it into my entry-level students. Because it just works. - *Tutor, Treatment group*

4.6.2 Shifting practice and embracing facilitation

Delivering the CA prompted some interviewed tutors to reflect on and adjust their teaching practices, with many embedding new principles into their FSQ level 1 classroom practice over time. Tutors found key aspects of the CA straightforward and readily adapted their existing teaching approaches to use visual tools (such as bar models and ratio tables). These strategies helped make abstract concepts more accessible, particularly for learners with dyslexia or processing differences. Several aspects of the approach required deliberate adjustment, including managing group dynamics, facilitating discussions, encouraging active listening and maintaining learner-led discussion. For some, this facilitative approach was unfamiliar.

> I have to physically, kind of, take myself back a little bit from, 'Right, this isn't me, this is going to be you,' and some groups have taken to that much better than others, but yes. For me, it is something I'm working on, not jumping straight into helping them - *Tutor, Treatment group*

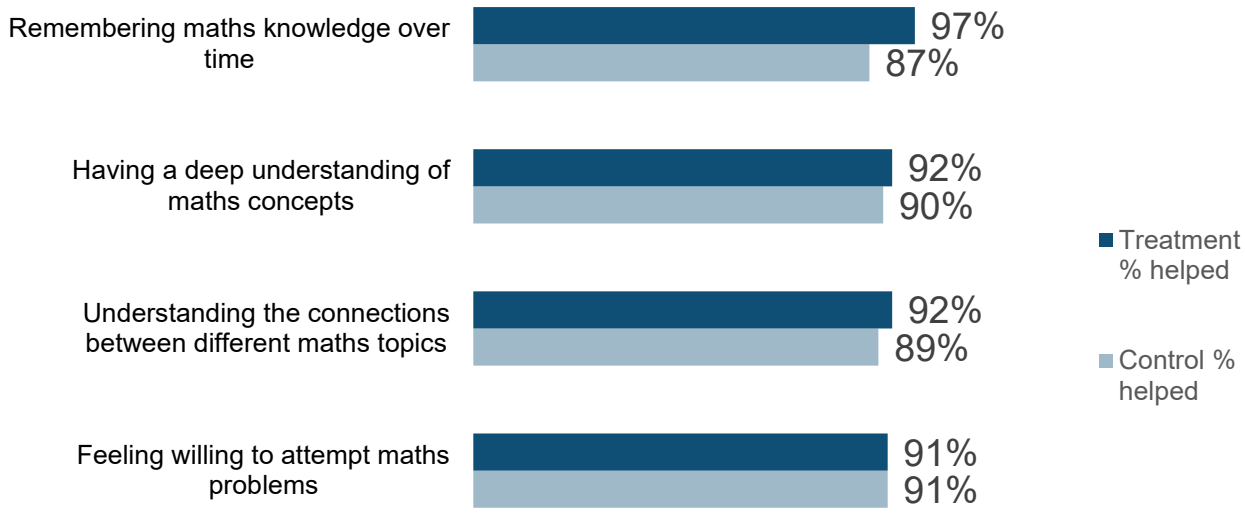
Trainers observed a positive shift and growing enthusiasm, with some tutors sharing learners' work and becoming more engaged in lesson study sessions as they saw improvements in learner self-efficacy and maths confidence. Interviewed tutors reported liking the facilitator role encouraged by the CA and noted increased learner engagement, confidence and ownership of problem-solving.

As teaching progressed, tutors observed further increases in learner engagement, development of critical thinking skills, a stronger sense of empowerment, and a deeper understanding of maths. They described how learning through exploration helped learners build confidence through ‘aha’ moments and develop greater ownership of their problem-solving. This was particularly helpful for adults who struggled with traditional methods, offering alternative ways to access and retain learning, whilst reducing the fear of making mistakes. This suggested that the mechanism of real-world contexts and learner-led problem-solving, anticipated by the theory of change was operating as planned.

4.6.3 Perceived impact on learners

In the endline survey, a range of questions were asked about learners' perceived impact of the course on various maths-related outcomes. **Error! Reference source not found.** shows the number of learners in the endline survey who reported strong perceived impacts on a range of maths outcomes related to the CA. The results indicate that self-reported perceived impacts were at similar levels across the treatment and control groups, except for 'remembering maths knowledge over time', where the proportion of learners who reported as the course as helpful was higher in the treatment group

Figure 6: What impact, if any, do you think the course has had on your …



(statistically significant at the 5% level).

Note: Results show what proportion of everyone who answered the question gave one of the positive response options ("Helped a lot", "Helped a little"). Results from the endline survey. Non- responses excluded. Missing N was 468 for all.

In the endline survey, questions were asked about learners' perceived impact on a range of wider outcomes. As shown in Table 25**Error! Reference source not found.**, among learners who responded, there were generally high impacts reported on mental health and wellbeing and sense of being part of a community. The proportion of learners who thought that the course had a positive impact on their chances of finding paid work, finding a more fulfilling job, or progressing / being promoted, was higher among the treatment group learners than among control group learners (significant at 5% level).

Table 25: What impact, if any, do you think the course has had on your …

Outcome	Control N	Control %	Treatment N	Treatment %
1.Mental health and wellbeing	121	76%	69	77%
2.Sense of being part of your local community	107	67%	65	72%
3.Chances of finding paid work in the future	119	74%	74	82%
4.Chances of earning a higher wage in future	63	77%	27	[u]
5.Chances of finding a job you find more fulfilling	53	65%	29	[u]
6.Chances of progressing / being promoted in your current job	46	56%	25	[u]

Note: Results show what proportion of everyone who answered the question gave one of the positive response options ("Very positive", "Fairly positive"). Results from the endline survey. Non-responses are excluded. Missing N was 469 for Q1, Q2, Q3, and 605 for Q4, Q5, Q5. Shorthand [u] indicates where percentages have been suppressed due to low reliability of cell counts (n < 30)

When asked about plans to do more courses or training, 86% of control group learners and 91% of treatment group learners responded ‘Yes - I plan to do more study or training’. Additionally, in a question relating specifically to whether the FSQ Level 1 course had affected participants’ plans to do more studying or training, 72% of the treatment group responded, “a lot” and 22% responded “a little”.

Both treatment and control group learners interviewed expressed similarly positive experiences. Among those experiencing CA, learners and tutors consistently noted that embedding maths in real-life contexts enhanced learners’ understanding of its relevance and practicality, contributing to more positive attitudes. They reported that this approach made learners more willing to attempt maths problems and use maths skills in daily life, as they could now apply them in real-life situations and had gained confidence through group discussions.

> I'm much more confident to solve the problems or to stop and try to think, ‘Let me see, I know how to do it, let me just take it easy and breath and think’… but I know how to start it, I know what to do. - *Learner, Treatment group*

Some treatment learners in interviews also referred to having increased self-efficacy and better skills for employment and life tasks. For example, they described applying their learning about percentages and fractions to work out shopping deals.

> I noticed that 'How can I work out these things? How can I find better deals?' and those things. You know, I'm a mum. We are always going shopping. These things were very helpful. - *Learner, Treatment group*

Other treatment learners said their skills would help them in their current or future jobs, and some of the learners looking for a job said they were confident their skills would help them get one.

> **Case vignette:** A learner with dyslexia whose job entailed work with sixth form SEND students, gained a greater understanding of maths language and structures aided by the visual approach. They used their new maths knowledge to support their students' maths learning and were looking to use their new FSQ qualification as a stepping stone to career progression in education. *Learner, Treatment group*

Interviewed treatment learners reported improved understanding of maths and connections between different maths topics. For example, they began to see how percentages and ratios were connected. They noted that a clear progression of topics, starting with simpler topics, helped them recognise these links. Their understanding was further reinforced through classwork, homework, and group activities, all of which provided opportunities to ask questions and consolidate their learning.

Control group learners also reported increased confidence, self-efficacy and maths skills in qualitative interviews. They gained an appreciation for the relevance of maths and its practical application to contexts such as shopping, budgeting, and home improvement, contributing to a more positive attitude towards maths. Some also felt more able to ask questions and persist with problems as the course progressed. They both valued their positive tutor relationships and peer support.

4.6.4 Perceived effects on learner engagement and confidence

Interviewed treatment tutors reported that learner-led and context-driven teaching fostered classroom cohesion and supported learner confidence. They described the learner-led approach as collaborative and inclusive, particularly in small groups, where it fostered a sense of cohesion and mutual support. Group discussions created a supportive environment that broadened learners' perspectives and demonstrated the real-world utility of maths, enhancing their confidence without the fear of making mistakes.

> Some of the questions are say what you see. They've been really useful in breaking down barriers in the class, of learners who don't like to speak out in that situation but felt confident to say something because there is no right or wrong answers…you can say whatever you want. It's building their confidence and that's really good. - *Tutor, Treatment group*

Some tutors reported strong engagement, particularly among learners with greater support needs or those returning to education after a long break. These learners often

responded well, especially if they had struggled with traditional maths methods. In contrast, more confident learners favoured familiar techniques and were less inclined to participate in group discussion or explore alternative strategies.

Some CA tutors said in interviews that engagement was more mixed among SEND and ESOL learners. Some SEND learners showed strong engagement, with the flexible, real-life context of the approach helping them connect with the material and build confidence. However, they said others found it challenging, especially those who preferred structured, rules-based learning or were less comfortable with group or discussion-based activities. Real-life scenarios helped some ESOL learners relate to the content and see its practical relevance. However, when examples were lengthy or linguistically demanding, they could exceed learners' language proficiency and become a barrier to understanding.

5 Conclusions

In concluding this report, it is crucial to acknowledge that trials of this type had not previously been delivered with adult learners in adult education settings. This trial demonstrates it is possible to lead intervention delivery and data collection in adult education settings, building confidence for future trialling of adult numeracy interventions.

5.1 Key findings

The CA intervention, underpinned by Realistic Mathematics Education (RME) was novel. While effectiveness had been established for school pupils in England by a team at Manchester Metropolitan University who developed it for use in Key stage 2 and early Key stage 3, it had not been tested with adults.

The impact evaluation found no statistically significant evidence that the CA improved the FSQ Level 1 maths pass rates of adult learners. The result was tested at the 95% confidence level, the conventional threshold used to assess if a finding is likely to reflect a genuine effect rather than chance and means that learners in the treatment group did no better or worse than those in control in terms of passing the FSQ tests.

When the analysis included only learners who met the attendance thresholds set by the product developer, the estimated impact on pass rates and robustness improved slightly, but results remained statistically insignificant at the 95% confidence level. Primary outcome data were missing for 38% of learners. Although this missingness was not predicted by treatment status or most baseline characteristics (except ethnicity and provider preference for CA in the trial gateway), it was strongly correlated with FSQ Level 1 maths course completion. When learners with missing primary outcomes were excluded, the treatment effect became weakly statistically significant (at the 90% confidence level rather than the standard 95% level). However, treatment status marginally predicted completion status, but the treatment effect was not statistically significant among compliers.

The findings for the effect of CA on maths confidence, a secondary measure for the trial, were inconclusive. These analyses were limited by low survey response rates, which means there is low confidence in these results. Two confidence measures were used. Each produced results in different directions, with one indicating a better outcome for the treatment group and the other showing a better outcome for the control group. It is not possible to say whether there is a promising effect on from the CA on adult learners' maths confidence.

The IPE established that the CA was delivered in line with product developer guidance. Tutors welcomed the professional development and refreshed approach to their teaching practice. Overall, 10 of 13 tutors who delivered the intervention attended all the pre-training and over 80% of the CPD sessions, the threshold set by the product developer for them to deliver with fidelity. There was evidence that tutors adapted training and

paced it differently than planned to accommodate some learner needs. This included providing catch-up sessions and reducing the pace of delivery. The extent of adaptations was limited, and changes were discussed with the product developer and judged to not affect the fidelity of CA delivery. The examples could be adopted into the intervention in future. Overall, the evidence suggests that the intervention is different from BAU teaching approaches, and the randomisation approach meant there was no overlap between the CA and BAU in the providers.

The CA was generally well received by learners who found the real-world application of maths engaging, noting it built their problem-solving capability and ability to understand maths in daily life. Attendance data showed that 59% of learners attended 10 of the 12 sessions i.e., more than 80% of sessions, which was the threshold for compliance in the impact analysis, the minimum level that product developers believed necessary for improved outcomes. Learners who attended fewer than 10 sessions cited personal and contextual barriers rather than course-related issues that may have affected pass rates regardless of the intervention.

5.2 Lessons for future delivery

Though well received, there were several challenges may explain why the positive experiences found in the IPE did not translate into measurable impacts. Key amongst these was the level of provider withdrawal following randomisation (42%). Of 50 providers recruited and randomised, 21 withdrew before or early in delivery. Main 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, or 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. This, combined with tutor comments on the intensity of training and CPD, may suggest potential issues for practicality that may require reconsideration. There were also requests from tutors for the CA to cover the FSQ Level 1 maths curriculum in full (15 sessions) rather than the 12 sessions it offered for the trial. For the trial, the RME approach was adapted based on a curriculum used with Key stage 2 and 3 school pupils, but this could be extended with newly developed content to cover the FSQ fully.

5.3 Considerations for future research

While the impact evaluation did not find statistically significant impacts, the pilot provides promising signs on the feasibility of conducting RCTs in adult learning settings and collecting the data necessary to understand any effect. It also demonstrates how the CA can be adapted to better accommodate diverse learner needs and tutor capacity. Most learners found the approach more meaningful that the traditional approach that leads with maths theory and considers real-world application only later. The intervention appears to be feasible for adult numeracy learners taking FSQ Level 1 maths. In any

future implementation the pilot evaluation evidence indicates additional guidance could be supplied on pacing and separately on support for ESOL learners. The training and CPD formats could be adapted to be more flexible to, for example, better accommodate part-time staff. Finally, the intervention could be extended to cover the FSQ Level 1 maths curriculum in full.

For future research, adaptations such as these should improve retention and fidelity, including reducing the training burden and making the course materials more appropriate to all students and delivery modes.

However, this pilot also reveals constraints for designing a fully-powered trial of CA in adult FSQ Level 1 courses. The number of adult education providers nationally, at c500 in any year, is small compared to the number of providers in other phases of education. Although all eligible providers were approached, only around 10% participated, and even fewer completed delivery. A fully-powered cluster-randomisation trial would require all eligible providers across England to sign up. Furthermore, pilot data on class numbers and sizes indicate that alternative randomisation strategies would not yield sufficient providers for a robust trial.

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).

References

Anghileri, J. (2001) *Development of division strategies for Year 5 pupils in ten English schools*. British Educational Research Journal, 27(1), pp. 85–103. Available at: https://doi.org/10.1080/01411920125581.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, *91*(434), 444-455.

Demack, S., Culliney, M., Boylan, M. and Wolstenholme, C. (2022) *Realistic Maths Education: Evaluation report*. Education Endowment Fund and Sheffield Hallam University.

Freudenthal, H. (1968) *Why to teach mathematics so as to be useful*. Educational Studies in Mathematics, 1(1), pp. 1–8. Available at: https://doi.org/10.1007/BF00426224.

Freudenthal, H. (1991) *Revisiting mathematics education: China lectures*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Hough, S., Solomon, Y., Dickinson, P. and Gough, S. (2017) *Investigating the impact of a Realistic Mathematics Education approach on achievement and attitudes in Post-16 GCSE resit classes*. Manchester Metropolitan University and The Nuffield Foundation.

Mathematics in Context (MiC) project (n.d.) *Mathematics in Context (MiC)*. [online]. (Institution/publisher details not provided – add if known). (Accessed: [add access date]).

Making Sense of Maths (2012) *RME textbook series, Making Sense of Maths*. (Publisher not provided – add if known).

Moffett, P. and Corcoran, D. (2010) *An evaluation of the implementation of Realistic Mathematics Education (RME) within primary schools in the North and South of Ireland: Final report*. Available at: https://scotens.org/site/wp-content/uploads/An-evaluation-of-the-implementation-of-Realistic-Mathematics-Education11.pdf (Accessed: 25 September 2025).

Realistic Mathematics Education (n.d.a) *How does RME do mathematics differently?* [online]. Available at: https://rme.org.uk/what-is-rme/about-rme/ (Accessed: 25 September 2025).

Realistic Mathematics Education (n.d.b) *Past projects*. [online]. Available at: https://rme.org.uk/about-us/past-projects/ (Accessed: 25 September 2025).

Searle, J. and Barmby, P. (2012) *Evaluation report on the Realistic Mathematics Education Pilot Project at Manchester Metropolitan University*. Centre for Evaluation and Monitoring (CEM), Durham University.

Ventistas, G., Ventista, O.M. and Tsani, P. (2024) *The impact of Realistic Mathematics Education on secondary school students' problem-solving skills: A comparative evaluation study*. Research in Mathematics Education, pp. 1–25. Available at: https://doi.org/10.1080/14794802.2024.2306633.

Zarnadze, B. (n.d.) *Mathematics in context: Do ratio tables help students make sense of maths problems?* Harlow College and Northampton College.

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 tutors 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.

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 2 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.

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.

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.

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.

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.

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 ongoing 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 collection 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.

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.

Power calculations: Statistical calculations performed before an RCT begins to ensure the study has sufficient statistical power to detect a significant effect. The primary purpose of a power calculation is to ensure a study is sufficiently "powered", meaning it has a high probability of finding a real effect. The calculation requires researchers to specify the desired level of statistical power, the significance level, and the smallest effect size they consider to be meaningful.

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.

Recall bias: A systematic error that occurs when people are asked to remember past events or experiences and their memories are incomplete or inaccurate.

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 three 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: The CA for FSQ Level 1 maths intervention content

About the intervention

The intervention integrates a Realistic Mathematics Education (RME) approach into the teaching of Functional Skills Maths curriculum for adult learners working towards Level 1. The approach is based on the principles of Realistic Maths Education (see www.rme.org.uk), which makes use of contexts, learner-driven models and a different role for the tutor in facilitating learner progress. The RME approach is different to standard approaches to learning maths in that it does not upfront formal methods. Adoption of these approaches has led to greater student engagement with mathematics, increased understanding of underlying concepts and improved problem-solving skills in several countries.

A critical aspect to the design is the way in which context is used. Contexts are not chosen because they might be motivating, nor are they chosen because they relate to another area of study. Instead, contexts are chosen because when “worked” they lead to models or representations that are useful for mathematical thinking.

Initial tutor training

Tutors had initial training amounting to 10 hours over 4 sessions, and ongoing weekly/fortnightly 2-hour online sessions provided by the CPD expert.

Tutors were required to listen to how their learners think and to work with them to develop and refine their informal strategies.

Teaching strategies aimed at promoting a different classroom culture, more learner led and facilitative, were highlighted in intervention training.

Course sessions

The lessons in this intervention were tailored for adult learners. The scheme consisted of 12 lessons, each of 1.5 hours duration.

Topics focussed on where maths ideas and strategies had potential for use beyond the 12 lessons. They were matched with Level 1 FSQ maths criteria, although Functional Skills was not covered in entirety. This meant that 50% of the FSQ Level 1 maths curriculum was covered in full, 25% was partially covered and 25% was not covered by the intervention.

The sequencing of the lessons was important and took account of the order in which various models are introduced, with the assumption that learners would have access to these models and strategies in subsequent lessons. The lessons as a whole placed an emphasis on proportional reasoning, which forms the bedrock of mathematical development at school level. Table A2 - 1 lists the modules, titles and key objectives for each of the 12 intervention delivery sessions.

Table A2 - 1: Overview of Contextualised Curriculum course sessions

Module	Lesson FS 1 content	Lesson objectives	Model or representation
Number 1 (4.5 hours)	Lesson 1: Percentage P1	• To build common-sense approaches to finding percentages • To develop imagery associated with landmark percentages of 100%, 50%, 25%, 75%, 10%, 1% • To develop a sense of proportion in relation to fractions and percentages • To use a percentage bar to find a percentage of an amount •	Percentage bar Matching and comparing Re-allotting (re-positioning) Array
	Lesson 2: Percentage P2	• To build common-sense approaches to finding percentages • To develop strategies for accurately positioning percentage amounts within a percentage bar • To draw and use a percentage bar to work out various percentage amounts • To notice that ***different*** routes on the percentage bar can lead to the ***same*** solution	
	Lesson 3: Percentage P3	• To build common-sense approaches to finding percentages • To extend use of a percentage bar to find percentage decrease and percentage increase • To recognise that ***different*** routes on the percentage bar can lead to the ***same*** solution • To develop calculator methods for finding percentages • To use the above approaches to answer exam questions	

Module	Lesson FS 1 content	Lesson objectives	Model or representation
Number 2 (3 hours)	Lesson 4: Ratio tables RT1	• To build common-sense approaches to scaling up and down quantities in proportion • To develop the use of the ratio table strategy as a way of working in proportion • To begin to recognise the range of questions for which a ratio table can be useful • To use ratio tables for exam-type questions	Ratio table and associated strategies (doubling, halving, adding, subtracting, spitting into three) Ratio table and anticipatory thinking
	Lesson 5: Ratio tables RT2	• To develop the use of the ratio table strategy as a way of calculating percentages. • To develop the use of the ratio table strategy to compare the size of fractions. • To use ratio tables in a wider range of proportional reasoning questions. • To use ratio tables for exam type questions.	
Data (3 hours)	Lesson 6: Data D1	• To build common-sense approaches for the need to group, organise and interpret data • To develop informal skills to divide up a number line as a scale • To develop the skills to draw bar charts, pie charts and line graphs • To use the above approaches to answer exam-style questions	Matching Bar model (whole numbers) Bar models (creating to scale) Re-allotting (levelling off bars) Positioning numbers on a scale
	Lesson 7: Data D2	• To build common-sense approaches to find mean and range • To develop the use of re-distributing and totalling up strategies as a way of finding the mean • To visualise the “gap” associated with the range • To use the above approaches to answer exam-style questions	

Module	Lesson FS 1 content	Lesson objectives	Model or representation
Geometry (4.5 hours)	Lesson 8: Geometry G1	• To build common-sense approaches to finding perimeters and to convert between units of length • To develop the use of "fitting in" strategies as a way of counting around a perimeter and of seeing one unit of length within another • To visualise perimeter in a variety of ways • To use the above approaches to answer exam-style questions	Matching and comparing Creating scales Measuring (repeated placing of informal units of length) Fitting in and filling Smart counting Ratio tables
	Lesson 9: Geometry G2	• To build common-sense approaches to finding area • To experience that *different* ways of counting tile pattern arrays can lead to the *same* totals • To practise scale drawing and fitting in a row of squares • To visualise a metre square and make meaning for "per square metre" • To use the above approaches to answer exam questions	
	Lesson 10: Geometry G3	• To experience that *different* ways of counting array patterns lead to the *same* totals • To connect square numbers with an array and with the ratio table • To connect "filling" with cubes and volume • To develop fast counting approaches for layers of cubes • To visualise a metre cube and a litre • To use the above approaches to answer exam questions •	

Module	Lesson FS 1 content	Lesson objectives	Model or representation
Number 3 (3 hours)	Lesson 11: Fractions	• To build on common-sense approaches to understanding fractions • To develop the use of the bar model as a way of working in fractions • To be able to calculate a fraction of an amount • To use the above approaches for exam-type questions	Fraction bar (fairly portioning, labelling each part, cumulative labelling) Drawing bespoke pictures Ratio table Ratio bar
	Lesson 12: Ratio	• To build common-sense approaches to finding, using and checking ratios • To continue to use the ratio table as a way of working with ratios • To visualise the parts of the bar associated with ratio • To use the above approaches to answer exam-style questions	

Source: Education and Training Foundation course materials

Appendix 3: Intervention cost data

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 is calculated as set out below.

Treatment group

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	The provider will receive a £1000 one off payment Tutors will receive 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	The provider will receive a £1000 one off payment.

The grant funding spend for the Adapted Mastery Approach and Contextualised Curriculum for FSQ Level 1 - Contextualised, based on the notes above, was £206,320. The training cost was £99,650 for Contextualised Curriculum and £106,870 for the Adapted Mastery Approach. This resulted in a total delivery spend for these interventions of £412,840.[14]

14 Figures rounded to the nearest £10

The product developer for the Adapted Mastery Approach and Contextualised Curriculum for FSQ Level 1 – Contextualised was the Education and Training Foundation (ETF). ETF's deliverables included development of all training materials in line with FSQ Level 1 Maths, deliver training to tutors, be available for feedback and questions, deliver the ongoing reflection sessions to tutors.

Appendix 4: Research questions and tools

IPE research questions in full

1. Has the intervention been delivered with fidelity (that is, in line with the intervention guidance)?
    a. Is the pre-delivery CPD training delivered as intended, and is it consistently delivered?
    b. How much of the pre-delivery CPD training do tutors attend?
    c. To what extent do tutors access the ongoing support throughout delivery of the intervention?
    d. To what extent do tutors implement the content, principles, approaches and activities in their teaching? How did they do this?
    e. 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?
    f. How feasible is it for tutors to implement the intervention as intended? Where this has not happened, what are the reasons for this? (the effects of tutors' initial and eventual maths confidence may be an interacting factor that will be explored as part of this).
2. To what extent, if at all, do the "business as usual" Functional Skills Level 1 courses typically include any content relating to the Contextualised Approach?
3. Is there any evidence of contamination of the "business as usual" Functional Skills 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 Functional Skills Maths to CA to FSQ Level 1 teaching practices? How does this vary between providers/tutors?
    b. To what extent, if at all, do Level 1 Function Skills tutors use CA content and principles in Function Skills Level 1 teaching?
        i. E.g., to what extent, if at all, do they use context to help explain maths concepts, and in what sequence with theoretical content are contexts used?
    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).
4. How did tutors experience delivering the intervention?

a. To what extent do tutors understand the rationale behind the intervention? To what extent did they agree with this rationale?
b. How do tutors feel about being asked to deliver the Contextualised Approach? What was the buy-in among tutors?
c. To what extent do tutors integrate CA into their lessons? To what extend did they make changes to their lesson plans? Did this change over time?
d. How do tutors feel about the integration of CA into their courses? How well does the format work?
e. How easy or difficult is it to deliver and integrate the interventions into the existing curriculum and course hours?
f. Are any of the principles, approaches and activities easier or harder to deliver than others?
g. To what extent is delivery substantially different to tutors' usual practices?
h. How did tutors feel about taking part in a trial? In particular, how did tutors delivering the "business as usual" approach feel about this?
i. What, if anything, do tutors feel they gained from the experience of delivering the intervention? How did these outcomes emerge and evolve?
    i. E.g. shift in teaching practices, shifting maths authority to learners (Contextualised Approach)
j. 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?
k. How, if at all, do tutors feel the intervention impact learners?
l. How, if at all, could the intervention be further developed or improved to enable wider rollout?

5. 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 intervention 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?

6. What are learners' experiences of the intervention?
    a. What motivates learners to take or re-take the Functional Skills Level 1 teaching course? What do they hope to gain?
    b. How engaged are learners with the course? How engaged are they in lessons and activities using the Contextualised Approach?
    c. What do learners think worked well and less well about the course? What could be improved?
    d. How familiar are the contexts provided by CA for FSQ Level 1 content, sessions and activities to learners? Does variance in levels of familiarity with context affect engagement, motivation or learning?
    e. Are learners able to move between informal contexts and abstract mathematical representations?
    f. Do learners gain an improved understanding of underlying mathematical concepts through the CA for FSQ Level 1 content, sessions and activities?
    g. Do learners become able to mathematise and use the language of maths more readily through CA for FSQ Level 1 content, sessions and activities?
7. What outcomes does the intervention have for learners?
    a. What, if anything, do learners feel they gained from the Contextualised Approach? (confidence, attitudes, willingness and skills to attempt maths, understanding of mathematical structure, self-efficacy, better skills for employment and life tasks).
        i. How does the intervention, and specifically the context-led approach, make a different to learners' perceived outcomes?
    b. Does the intervention create interest in taking up a further maths course?
    c. Does the intervention lead to unplanned/unanticipated outcomes?
    d. Do learners gain a greater willingness to try different types of questions and problem-solving tasks as a result of the intervention?
    e. What elements were most useful, and were perceived to make the most difference?
    f. To what extent do learner outcomes vary by individual characteristics?
8. What enablers and barriers are there to learner engagement and participation in the intervention?
9. 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?

10. 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?
    h. Were the outcome measures used valid, reliable, acceptable and complete? How / in what ways could they be improved?
    i. 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?

5.4 IPE survey

Multiply learner endline questionnaire – FSQ V1

KEY

QUESTION NAME IN BOLD
ROUTING / SCRIPTING INSTRUCTIONS IN PURPLE
ANSWER CODE SCRIPTING INSTRUCTIONS IN ORANGE
CATI ONLY TEXT IN BLUE HIGHLIGHT
CAWI ONLY TEXT IN YELLOW HIGHLIGHT

Trial variables

DO NOT SHOW TO RESPONDENTS / FROM SAMPLE
TRIALDUM – Trial enrolled in

1. Embedded Maths
2. Maths FSQ Trial
3. Preparation for Maths GCSE
4. Family Numeracy

DO NOT SHOW TO RESPONDENTS / FROM SAMPLE
GROUPDUM – Trial group

1. Treatment
2. Treatment – Maths Mastery
3. Treatment – Contextualised Curriculum
4. Control

DO NOT SHOW TO RESPONDENTS / FROM SAMPLE
TIMEDUM – Baseline or endline survey

1. Baseline
2. Endline

CAWI Intro

SHOW IF CAWI

INTRO_CAWI Welcome to this survey which is being carried out by Ipsos on behalf of the Department for Education.

You are being asked to complete this survey because you have taken part in Functional Skills Maths Level 1 course. You may have been asked to complete a similar survey around the start of the course. We are getting in touch to find out if anything has changed for you since starting the course and to hear your views about the course. Your answers will help the Department for Education to improve learning opportunities in the future.

The survey should take about 20 minutes to complete. Taking part in the survey is voluntary and you don't have to answer any questions that you don't want to. The information you provide will be used for research purposes only and it will not be possible to identify you in the research findings.

If you have any queries, please contact the Ipsos research team by emailing EducationTrials@Ipsos.com or calling us for free on 0800 470 2983 and leaving a message with your name and phone number so we can give you a call back.

To find out more, and view the privacy notice, click here: [privacy notice]

To begin the survey, please click next.

CAWI ONLY
ASK ALL / TEXT SCREEN
QINTPI Firstly we would like to ask you for your name, date of birth and postcode. This is so that we can match your answers with the details you provided when you started the course.

CAWI ONLY
ASK ALL / SINGLE CODE
QFIRSTNAME What is your first name?
1. Please write in [OPEN]
2. Prefer not to say

CAWI ONLY
ASK ALL / SINGLE CODE
QSURNAME And what is your surname?
1. Please write in [OPEN]
2. Prefer not to say

CAWI ONLY
ASK ALL / SINGLE CODE
QDOB What is your date of birth?
1. [SHOW AS 00/00/0000]
2. Prefer not to say

CAWI ONLY
ASK ALL / SINGLE CODE
QPOSTCODE What is your postcode?

1. Please write in [OPEN]
2. Prefer not to say

CATI – Intro / Screener

SHOW IF CATI
INTRO_CATI Good morning / afternoon / evening, my name is [name] and I am calling from Ipsos.

Please can I speak to [INSERT NAME FROM SAMPLE]?

We have been commissioned by the Department for Education to carry out some research on what works to improve adults' numeracy skills. [INSERT NAME FROM SAMPLE] is being asked to complete this survey because they have taken part in a Functional Skills Maths Level 1 course.
The interview will take about 30 minutes and will ask about their experiences of the course.

IF NECESSARY: You can access the privacy notice for this survey online at [privacy notice url]

Is now a good time to speak?

CATI ONLY
ASK ALL
SINGLE CODE
ONLINEREQUEST Thank you for speaking with me. As mentioned, we are contacting you because you have taken part in a course. You may have also taken part in a survey nearer the start of your course.

As part of this research, we are conducting a follow-up survey now that your course has come to an end or is about to come to an end. Your answers will help the Department for Education to improve opportunities like the one you signed up to, and other learning opportunities, in the future.

Some people may have already been asked to complete this survey online by their course tutors or administrators. Please can I check, were you asked to complete this follow-up survey online?

1. Yes
2. No [GO TO CATICONSENT]
3. Don't know / can't remember [GO TO CATICONSENT]
4. DO NOT READ OUT[15] Prefer not to say [GO TO CATICONSENT]

CATI ONLY
ASK IF ONLINEREQUEST = 1 / SINGLE CODE

[15] CATI only

ONLINECOMPLETE And to what extent did you complete the follow-up survey online, if at all?

1. Completed the follow-up survey online
2. Started the follow-up survey online but didn't complete it [GO TO CATICONSENT]
3. Couldn't complete the follow-up survey online [GO TO CATICONSENT]
4. Refused to complete the follow-up survey online [GO TO CATICONSENT]
5. Don't know / can't remember [GO TO CATICONSENT]
6. DO NOT READ OUT[16] Prefer not to say [GO TO CATICONSENT]

CATI ONLY
ASK IF ONLINECOMPLETE = 1
ONLINECOMPLETE_TIME Thank you for confirming. How long ago did you complete the follow-up survey online?

PROMPT AS NECESSARY[17]

1. Today
2. Yesterday
3. 2 days ago
4. 3 days ago
5. More than 3 days ago [GO TO CATICONSENT]
6. Don't know / can't remember [GO TO CATICONSENT]
7. DO NOT READ OUT[18] Prefer not to say [GO TO CATICONSENT]

CATI ONLY
IF ONLINECOMPLETE_TIME = 1, 2, 3, 4
ONLINECHECK Thank you for letting us know. We will check our records again. Thank you for your time. We may give you a call back if we find that we do not have a follow-up survey record for you. [END INTERVIEW][19]

CATI ONLY
ASK ALL / SINGLE CODE
CATICONSENT Thank you for letting us know. We would be very grateful if you could take part in a telephone interview to complete the follow-up survey now that your course is coming to an end or has now ended.

[REPLACE TEXT IF ONLINECOMPLETE = 2 Thank you for letting us know. We would be very grateful if you could take part in a telephone interview to complete the follow-up survey now that your course is coming to an end or has now ended. Unfortunately, we will need to start the survey again from the beginning. We're very sorry for the inconvenience.]

[REPLACE TEXT IF ONLINECOMPLETE_TIME = 5, 6, 7 Thank you for letting us know. Unfortunately, this means that we do not have a complete survey record for you. We would be very grateful if you could take part in a telephone interview to complete the follow-up survey now that your course is coming to an end or has now ended. We're very sorry for the inconvenience.]

[16] CATI only
[17] CATI only
[18] CATI only
[19] CATI only

IF NECESSARY: Anything you tell us will be held in the strictest confidence. The information you provide will be used only for research purposes and the research findings will not identify you.

Are you happy to proceed with the interview?

1. Yes
- No [END INTERVIEW OR MAKE HARD APPOINTMENT][20]
-

CATI ONLY
ASK ALL
QCONFIRM Firstly, I would like to check that we have the correct details about you so that we can match your answers with the details you provided when you started the course.

CATI ONLY
ASK IF FIRSTNAME IN SAMPLE IS COMPLETE
QCFIRSTNAME Can I confirm your first name is…?

SHOW FIRSTNAME FROM SAMPLE

PROMPT AS NECESSARY[21]
1. Yes, that is correct
2. No that is not correct (please specify) [OPEN] INPUT UPDATE[22]
3. DO NOT READ OUT[23] Prefer not to say

CATI ONLY
ASK IF QCFIRSTNAME = 2 OR FIRSTNAME IN SAMPLE IS BLANK
QFIRSTNAME Can I take your first name?
1. Please write in [OPEN]
2. DO NOT READ OUT[24] Prefer not to say

CATI ONLY
ASK IF SURNAME FROM SAMPLE IS COMPLETE
QCSURNAME
And your surname is…?

SHOW SURNAME FROM SAMPLE

PROMPT AS NECESSARY[25]
1. Yes, that is correct
2. No that is not correct (please specify) [OPEN] INPUT UPDATE[26]
3. DO NOT READ OUT[27] Prefer not to say

[20] CATI only
[21] CATI only
[22] CATI only
[23] CATI only
[24] CATI only
[25] CATI only
[26] CATI only
[27] CATI only

CATI ONLY
ASK IF QCSURNAME = 2 OR SURNAME IN SAMPLE IS BLANK / SINGLE CODE
QSURNAME Can I take your surname?

1. Please write in [OPEN]
2. DO NOT READ OUT[28] Prefer not to say

CATI ONLY
ASK IF DOB IN SAMPLE IS COMPLETE
QCDOB Can I confirm your date of birth is…?

SHOW DOB FROM SAMPLE

PROMPT AS NECESSARY[29]

1. Yes, that is correct
2. No that is not correct (please specify) [OPEN] INPUT UPDATE[30]
3. DO NOT READ OUT[31] Prefer not to say

CATI ONLY
ASK IF QCDOB = 2 OR DOB IN SAMPLE IS BLANK / SINGLE CODE
QDOB Can I take your date of birth?

1. [SHOW AS 00/00/0000]
2. DO NOT READ OUT[32] Prefer not to say

CATI ONLY
ASK IF POSTCODE IN SAMPLE IS COMPLETE
QCPOSTCODE Can I confirm your postcode is…?

SHOW POSTCODE FROM SAMPLE

PROMPT AS NECESSARY[33]

1. Yes, that is correct
2. No that is not correct (please specify) [OPEN] INPUT UPDATE[34]
3. DO NOT READ OUT[35] Prefer not to say

CATI ONLY
ASK IF QCPOSTCODE = 2 OR POSTCODE IN SAMPLE IS BLANK / SINGLE CODE
QPOSTCODE Can I take your postcode?

1. Please write in [OPEN]
2. DO NOT READ OUT[36] Prefer not to say

[28] CATI only
[29] CATI only
[30] CATI only
[31] CATI only
[32] CATI only
[33] CATI only
[34] CATI only
[35] CATI only
[36] CATI only

Course confirmation

ASK IF TRIALDUM = 1-3 / SINGLE CODE

QMULTIP We understand that you have completed, or are about to complete, the following course:

[SHOW IF TRIALDUM = 1] Health & Social Care Level 2
[SHOW IF TRIALDUM = 2] Functional Skills Maths Level 1
[SHOW IF TRIALDUM = 3] Maths GCSE

Is this correct?

PROMPT AS NECESSARY[37]

1. Yes, this was the correct course
2. No, it was a different course [SCREEN OUT]

QCOMPLETE And did you complete this course?

IF NECESSARY[38] If you attended the whole course but didn't pass, please answer 'Yes – completed course'.

1. Yes – completed course
2. No – left course part way through
3. No – course is still in progress
4. DO NOT READ OUT[39] Don't know
5. DO NOT READ OUT[40] Prefer not to say

ASK IF QCOMPLETE = 2 'NO – LEFT COURSE PART WAY THROUGH' / MULTI CODE / RANDOMISE CODES 1-12

QINCOMPLETEALL Why didn't you stay to the end of the course?

Please select all that apply[41]

READ OUT, SELECT ALL THAT APPLY[42]

1. Concerns about the cost of attending
2. Didn't like what I was being taught
3. Did not like the way the course was delivered (online or in a classroom)
4. Didn't like the other people on the course
5. It was too difficult
6. It was too easy
7. Not enough time or conflicted with other commitments

[37] CATI only
[38] CATI only
[39] CATI only
[40] CATI only
[41] CAWI only
[42] CATI only

8. Lack of support from family
9. Lack of support from employer
10. Personal / family problems / ill health
11. Course no longer relevant to my job
12. Did not like tutor / teacher
13. Another reason (please type in) _____[OPEN]
14. DO NOT READ OUT[43] Don't know [EXCLUSIVE]

Opening demographics

ASK ALL / TEXT SCREEN
QINTROWORK We would like to ask you a couple of questions about yourself to see if anything has changed since you started the course. Your response to these questions will be used for research purposes only.

ASK ALL / SINGLE CODE

QCURRENTWORK Which of the following best describes the main thing you are currently doing now the course is coming to an end or has ended?

IF NECESSARY[44] If you are doing more than one thing, please select the one you spend the most hours doing.

READ OUT, SINGLE CODE[45]

1. Working as an employee part-time (less than 30 hours per week)
2. Working as an employee full-time (more than 30 hours per week)
3. Self-employed or freelance
4. Temporarily away from work ill, on holiday or temporarily laid off
5. On maternity or paternity leave
6. Doing any other kind of paid work
7. Retired (whether receiving pension or not)
8. Studying
9. Looking after home or family
10. Long-term sick or disabled
11. Waiting to start a job already accepted
12. Unemployed and looking for work
13. Unemployed and not looking for work
14. Something else (please type in) [OPEN]
15. DO NOT READ OUT[46] Prefer not to say

[43] CATI only
[44] CATI only
[45] CATI only
[46] CATI only

QPARENT How many children under the age of 18 do you have parental responsibility for and/or legal guardianship?

By parental responsibility we mean providing a home and protecting and maintaining a child.

1. Numeric [1-10]
2. None [EXCLUSIVE]
- DO NOT READ OUT[47] Prefer not to say [EXCLUSIVE]

Working with numbers

ASK ALL / TEXT
QINTRO1 The next few questions are about how you feel about working with numbers now that your course is coming to an end or has now ended.

ASK ALL / SINGLE CODE
QCONF 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

ASK ALL / SINGLE CODE / RANDOMISE STATEMENTS
QNUMCONF And now how confident do you feel about doing the following things in everyday life?

READ OUT, SINGLE CODE PER STATEMENT[48]

a) Checking your change is right when you have bought something
b) Working out the best deals when shopping
c) IF HAS PARENTAL RESPONSIBILITIES QPARENT CODE 1=1-10 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 QCURRENTWORK =1-6 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

[47] CATI only
[48] CATI only

ASK ALL / SINGLE CODE / RANDOMISE STATEMENTS
QNUMBHELP To what extent, if at all, do you think this course has helped to improve your confidence when doing the following things in everyday life?

READ OUT, SINGLE CODE PER STATEMENT[49]

a) Checking your change is right when you have bought something
b) Working out the best deals when shopping
c) [IF HAS PARENTAL RESPONSIBILITIES QPARENT CODE 1=1-10] Helping children with 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 QCURRENTWORK =1-6] Working with numbers as part of a job

1. Helped a lot
2. Helped a little
3. Made no difference
4. Made it a little worse
5. Made it a lot worse
6. Not relevant to me
7. DO NOT READ OUT[50] Don't know

ASK ALL / SINGLE CODE / RANDOMISE STATEMENTS
QMATHSHELP What impact, if any, do you think the course has had on your…

READ OUT, SINGLE CODE PER STATEMENT[51]

a) Understanding maths structures
b) Understanding the connections between different maths topics
c) Feeling willing to attempt maths problems
d) Having a deep understanding of maths concepts
e) Remembering maths knowledge over time

1. Helped a lot
2. Helped a little
3. Made no difference
4. Made it a little worse
5. Made it a lot worse
6. DO NOT READ OUT[52] Don't know

ASK ALL / SINGLE CODE PER STATEMENT / RANDOMISE STATEMENTS
QLEARN2 How do you feel about each of the following?

READ OUT, SINGLE CODE PER STATEMENT[53]

a) Doing more learning, on a range of subjects
b) Doing more maths learning

[49] CATI only
[50] CATI only
[51] CATI only
[52] CATI only
[53] CATI only

c) Getting more qualifications

1. Much more interested
2. A little more interested
3. No change in interest
4. A little less interested
5. A lot less interested
6. DO NOT READ OUT[54] Don't know

ASK ALL / SINGLE CODE PER STATEMENT / RANDOMISE STATEMENTS

QBEN What impact, if any, do you think the course has had on your …

READ OUT, SINGLE CODE PER STATEMENT[55]

a) Mental health and wellbeing
b) Sense of being part of your local community
c) Chances of finding paid work in the future
d) [IF WORKING (QCURRWORK = 1-5)] Chances of earning a higher wage in future
e) [IF WORKING (QCURRWORK = 1-5)] Chances of finding a job you find more fulfilling
f) [IF WORKING (QCURRWORK = 1-5)] Chances of progressing / being promoted in your current job

1. Very positive
2. Fairly positive
3. No impact
4. Fairly negative
5. Very negative

- DO NOT READ OUT[56] Don't know
-

ASK IF QBEN = 4 - 6 AT ALL STATEMENTS SHOWN (NO BENEFIT AT ALL STATEMENTS) / MULTI CODE / RANDOMISE CODES 1-7

QNOBEN Why do you think the course hasn't helped with any of these things?

Please select all that apply[57]

*REA*D OUT, SELECT ALL THAT APPLY[58]

[54] CATI only

[57] CAWI only
[58] CATI only

1. My maths / numeracy was already good enough
2. It's too soon after the course to say
3. Quality of teaching / materials available wasn't good enough
4. What I learned was not helpful or relevant to my life
5. It was too easy / I didn't learn anything new
6. It was too hard
7. I need a more advanced course / further training to see any benefits
8. Another reason (please type in) _____[OPEN]
9. DO NOT READ OUT[59] Don't know [EXCLUSIVE]

Course information

ASK IF ALL / SINGLE CODE

QDELIV The next few questions are about how the course was delivered.

How was the majority of your course delivered?

READ OUT, SINGLE CODE[60]

1. 'In person' only. By in-person we mean in the same room as the teacher
2. Online only. By online we mean learning delivered by your teacher online, e.g. via Zoom, pre-recorded sessions with teachers giving a lesson, and completing quizzes and exercises that the computer 'marks' and sets a next question for you
3. A mix of online and 'in person' teaching

•

ASK ALL / SINGLE CODE PER STATEMENT / RANDOMISE STATEMENTS

QELEM As part of your course, did you have access to any of the following?

READ OUT, SINGLE CODE PER STATEMENT[61]

a) Teaching delivered online, either via video or live e.g. by Zoom
b) Online resources such as learning materials, workbooks, reading materials
c) A teacher you could see in person
d) A teacher you could contact by telephone, email or online via e.g. Zoom or a chat function
e) A peer support network organised by the learning provider where learners share knowledge, experience, or practical help with each other
f) Online quizzes and exercises that the computer 'marks'

1. Yes – I used this
2. Yes – but I did not use this
3. No – it was not offered
4. DO NOT READ OUT[62] Don't know

ASK ALL / SINGLE CODE PER STATEMENT / RANDOMISE STATEMENTS

[59] CATI only
[60] CATI only
[61] CATI only
[62] CATI only

QAPPROACH Did your tutor use any of the following approaches when delivering your course?

READ OUT, SINGLE CODE PER STATEMENT[63]

a) Using real life examples to explain maths problems
b) Encouraging you to use problem solving strategies to think through and respond to maths problems
c) Getting students to think through maths problems together
d) Encouraging you to listen to how other students solve problems
e) Using pictures and physical materials to make maths easier to understand
f) Giving you enough time to fully understand a topic before moving on
g) Breaking down ideas into small steps to show how they fit together and build into the larger concept and across the curriculum
h) Showing the same idea in different ways to help you understand underlying maths principles
i) Valuing your existing maths knowledge and helping you build on it
j) Helping you see how different maths topics connect
k) Spending time understanding patterns and relationships between maths concepts
l) Encouraging you to use clear mathematical language to explain your thinking
m) Developing a culture in which everyone believes they can succeed

1. Yes – A lot
2. Yes – A little
3. No
4. DO NOT READ OUT[64] Don't know

ASK IF 'YES' AT STATEMENTS IN QAPPROACH [QAPPROACH a-m = 1 OR 2] / SINGLE CODE

QAPPROACHUSE How helpful did you find each of these approaches

SHOW STATEMENTS A-M AS ROWS WHERE CODES 1 OR 2 SELECTED

READ OUT, SINGLE CODE PER STATEMENT[65]

1. Very helpful
2. Somewhat helpful
3. Neither helpful not unhelpful
4. Somewhat unhelpful
5. Very unhelpful

- DO NOT READ OUT[66] Don't know

Your experience on the course / programme

ASK ALL / SINGLE CODE PER STATEMENT / RANDOMISE STATEMENTS

[63] CATI only
[64] CATI only
[65] CATI only
[66] CATI only

QEXPEC3 How did each of the following compare with your expectations when you signed up for the course?

READ OUT, SINGLE CODE PER STATEMENT[67]

a) [IF RECEIVED IN PERSON TEACHING, QDELIV=1 OR 3] The quality of teaching you received in person as part of this course
b) [IF RECEIVED ONLINE TEACHING, QDELIV=2 OR 3] The quality of teaching delivered online, this includes both learning delivered by your teacher online, e.g. via Zoom, pre-recorded sessions with teachers giving a lesson and completing quizzes and exercises that the computer 'marks' and sets a next question for you
c) The communication you received throughout the course
d) The content of any learning materials, workbooks or reading materials
e) The length of each lesson or session
f) The overall length of your course
g) The overall learning experience

1. Much better than expected
2. A little better than expected
3. About the same as expected
4. A little worse than expected
5. Much worse than expected

- DO NOT READ OUT[68] Don't know / can't remember

ASK ALL / SINGLE CODE

QEASE Would you say that the maths covered on your course was…?

READ OUT, SINGLE CODE[69]

1. Too easy
2. About right
3. Too difficult
4. DO NOT READ OUT[70] Don't know / can't remember

Future plans

ASK ALL / SINGLE CODE

QTRAINFO Have you received any information about other courses you could do after this course?

1. Yes
2. No

- DO NOT READ OUT[71] Don't know / can't remember

ASK ALL / SINGLE CODE

[67] CATI only
[68] CATI only
[69] CATI only
[70] CATI only
[71] CATI only

QTRAIN1 Are you planning to do any more courses or training?

PROMPT AS NECESSARY[72]

1. Yes – I plan to do more study or training
2. No
3. DO NOT READ OUT[73] Prefer not to say

ASK IF PLANNING TO DO MORE STUDYING (QTRAIN1=1) / SINGLE CODE

QTRAIN2 What sort of courses or training are you planning to do?

[If you are planning on doing more than one course, please think about the one you will start first.][74]

READ OUT, SINGLE CODE[75]

1. An apprenticeship
2. [IF WORKING (QCURRWORK = 1-5)] Course or training at work, or from your employer
3. Course or training related to your interests, hobbies or for personal development
4. Other course or training at a further education (FE) /adult education college or training provider
5. Other course or training at a higher education institution, e.g. university
6. Another type of course or training – Please type in [OPEN]
7. DO NOT READ OUT[76] Prefer not to say

ASK IF PLANNING TO DO MORE STUDYING (QTRAIN1=1)/ SINGLE CODE

QTRAIN4 To what extent, if any, did your experience of the course influence your decision to do more study or training?

READ OUT, SINGLE CODE[77]

1. A lot
2. A little
3. Not at all
4. DO NOT READ OUT[78] Don't know

Overall satisfaction

ASK ALL / TEXT

QINTROSAT The next few questions are about what you thought about your course.

ASK ALL / SINGLE CODE

[72] CATI only
[73] CATI only
[74] If Qtrain1=1 show 'If you are doing more than one course, please think about the one you started first.'. If Qtrain1=2 show 'If you are planning on doing more than one course, please think about the one you will start first.'
[75] CATI only
[76] CATI only
[77] CATI only
[78] CATI only

QSATALL Overall, how satisfied or dissatisfied were you with the course?

READ OUT, SINGLE CODE[79]

1. Very satisfied
2. Fairly satisfied
3. Neither satisfied nor dissatisfied
4. Fairly dissatisfied
5. Very dissatisfied
6. DO NOT READ OUT[80] Don't know

ASK IF DISSATISFIED (QSATALL=4-5) / MULTI CODE / RANDOMISE CODES 1-13

QDISAT Why were you dissatisfied with the course overall?

Please select all that apply[81]

READ OUT, SELECT ALL THAT APPLY[82]

1. It was too difficult
2. It was too easy
3. Poor quality of teaching
4. Less contact time with teachers than expected
5. The classes were too long/took too much time
6. [IF IN PERSON (QDELIV = 1 OR 3)] Difficult to travel to the lessons
7. Problems with the technology
8. Lessons were at an inconvenient time
9. Didn't like the other people on the course
10. Difficult to juggle studying with my other commitments
11. Did not like the tutor/teacher
12. Did not like the way the course was taught (online or in a classroom)
13. It made my anxiety about maths worse
14. Another reason (please type in) _____[OPEN]
15. DO NOT READ OUT[83] Don't know [EXCLUSIVE]

ASK ALL / SINGLE CODE PER STATEMENT/ RANDOMISE STATEMENTS

QSATOTH How satisfied or dissatisfied were you with each of the following?

READ OUT, SINGLE CODE PER STATEMENT[84]

a) [IF RECEIVED IN PERSON TEACHING (QDELIV=1 OR 3)] The quality of teaching you received in person as part of this course
b) [IF RECEIVED ONLINE TEACHING (QDELIV=2 OR 3)] The quality of teaching delivered online, either via video or live e.g. by Zoom, this includes both learning delivered by your teacher online, eg via Zoom, pre-recorded sessions with teachers giving a lesson and completing quizzes and exercises that the computer 'marks' and sets a next question for you

[79] CATI only
[80] CATI only
[81] CAWI only
[82] CATI only
[83] CATI only
[84] CATI only

c) The content of any learning materials, workbooks or reading materials
d) The length of each lesson or session
e) The overall length of your course
f) The time of day the lessons were held
g) [IF RECEIVED INFORMATION ABOUT OTHER COURSES (QTRAINFO=1)] The information provided about other courses to do
h) The difficulty of the maths covered
i) The overall learning experience

1. Very satisfied
2. Fairly satisfied
3. Neither satisfied nor dissatisfied
4. Fairly dissatisfied
5. Very dissatisfied
6. Don't know

ASK IF DISSATSIFED WITH LEARNING MATERIALS (QSATOTH= 4 or 5 at code C) / MULTI CODE / RANDOMISE CODES 1-6

QDISWHYCONT You said you were dissatisfied with the content of the learning materials. Why was this?

Please select all that apply[85]

READ OUT, SELECT ALL THAT APPLY[86]

1. It was too easy
2. It was too difficult
3. It took too long to complete
4. Instructions were unclear
5. Not a good mix of theory (e.g. watching the teacher show us how to use multiplication, division, maths formulas etc.) and practical (e.g. working through maths questions myself)
6. Problems with the technology / the technology did not work
7. Another reason (please type in) _____[OPEN]
8. DO NOT READ OUT[87] Don't know [EXCLUSIVE]

ASK IF DISSATISFIED ABOUT INFORMATION ABOUT OTHER COURSES (QSATOTH= 4 OR 5 AT CODE G) / MULTI CODE / RANDOMISE CODES 1-6

QDISWHYINFO You said you were dissatisfied with the information provided about other courses to do after this one. Why was this?

Please select all that apply[88]

[85] CAWI only
[86] CATI only
[87] CATI only
[88] CAWI only

READ OUT, SELECT ALL THAT APPLY[89]

1. Did not understand the information
2. Did not receive enough information
3. Information was too general
4. Information was too specific
5. Courses were not appropriate for me
6. Not interested in doing further courses
7. Another reason (please type in) _____[OPEN]
8. DO NOT READ OUT[90] Don't know [EXCLUSIVE]

ASK IF DISSATISFIED WITH QUALITY OF TEACHING (QSATOTH= 4 OR 5 AT CODE A OR B) / MULTI CODE / RANDOMISE CODES 1-7

QDISWHYTEACH You said you were dissatisfied with the quality of the teaching. Why was this?

Please select all that apply[91]

READ OUT, SELECT ALL THAT APPLY[92]

1. I did not like the style of teaching
2. Not enough interaction in the lessons
3. The teacher did not explain things clearly / give clear instructions
4. Not a good mix of theory (e.g. watching the teacher show us how to use multiplication, division, maths formulas etc.) and practical (e.g. working through maths questions myself)
5. The teacher was late to lessons
6. The teacher was not able to control the class/prevent disruptions
7. Not enough time spent with the teacher
8. Another reason (please type in) _____[OPEN]
9. DO NOT READ OUT[93] Don't know [EXCLUSIVE]

ASK ALL / SINGLE CODE

QENGAGING Overall, how engaging did you find the course?

1. Very engaging
2. Fairly engaging
3. Neutral
4. Not very engaging
5. Not engaging at all

- DO NOT READ OUT[94] Don't know
-

[89] CATI only
[90] CATI only
[91] CAWI only
[92] CATI only
[93] CATI only
[94] CATI only

ASK ALL / SINGLE CODE

QREC How likely are you to recommend a similar course to people you know?

READ OUT, SINGLE CODE[95]

1. I have already recommended it
2. Very likely
3. Quite likely
4. Neither likely nor unlikely
5. Quite unlikely
6. Very unlikely
7. DO NOT READ OUT[96] Don't know

ASK ALL / OPEN

QCHANGE. Thinking about your whole experience of your course, what would be one thing that you would change to improve it for other learners?

1. OPEN
2. Nothing
3. DO NOT READ OUT[97] Don't know

Recontact

ASK ALL / SINGLE CODE
QRECONTACT Ipsos UK would like to speak to some learners again for a longer 45-minute interview to discuss their experiences of completing the course. Would you be happy for someone from Ipsos to contact you and invite you to participate in an interview?

All of our research is voluntary, so if you say yes now, you don't have to take part when invited.

1. Yes
2. No

- DO NOT READ OUT[98] Don't know
- DO NOT READ OUT[99] Prefer not to say

ASK IF QRECONTACT=1 (HAPPY TO BE RECONTACTED) / OPEN
QRECONTACTEMAIL Please provide an email address that you would be happy to be contacted on.

Email address:

Confirm email address:

DO NOT READ OUT[100] Prefer not to say

---

[95] CATI only
[96] CATI only
[97] CATI only
[98] CATI only
[99] CATI only
[100] CATI only

ASK IF QRECONTACT=1 (HAPPY TO BE RECONTACTED) / OPEN
QRECONTACTTEL Please provide a telephone number that you would be happy to be contacted on.

Telephone number

Confirm telephone number:

DO NOT READ OUT[101] Prefer not to say

Survey end

ASK ALL / TEXT SCREEN
QEND This is the end of the survey.

Thank you for taking the time to complete this important survey. Your contributions have provided an invaluable resource to help the Department for Education and we are very grateful for your time.

To get more information about the research you can contact us using the details below:

Email: EducationTrials@Ipsos.com

IPE topic guides

Wave 1: Tutor Interview Topic Guide – Treatment (Contextualised /Mastery)

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

I work for IES, an independent, not-for-profit research organisation, and King's College London, a public research university and higher education institution. Along with Ipsos and King's College London/Institute for Employment Studies, our research team has been commissioned by the Department of Education to evaluate the Contextualised Maths/Adapted 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 Contextualised maths/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.

[101] CATI only

Today's discussion will cover your experiences of Contextualised Maths/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 Contextualised/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 Contextualised/Maths Mastery approaches and principles?
6. How different, if at all, are 'normal' teaching practices for functional skills maths compared to Contextualised/Maths Mastery approaches and principles?

Researcher note:

Professional development training includes 10 hours spread over 4 sessions. The training covers the Contextualised Approach, its rationale, and the scheme of work. There are 12 lesson plans for Contextualised and 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.

Section B: Training and Resources

7. We understand you attended [use sampling spreadsheet to determine attendance] amount of training. Is this correct?
    a. *[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?
    b. *[If they attended some training sessions or none]*: How was information for missed sessions shared with you, if at all?
8. What did you think of the training sessions you attended? Probe for content, format, and interaction etc
9. What was most useful about the training? What was least useful?
10. Is there anything you think could be changed or improved about the training?
11. *Did you feel the training gave you a good understanding of the Contextualised/Maths Mastery approach and teaching strategies? *If not,* which aspects were you less certain about?*

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 trainer. 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.

12. 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?*
13. 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*
14. To what extent are you finding these session(s) useful? Why/why not?

Other support

15. Have you had any contact with the training team at ETF for additional support outside the training or lesson study sessions?
    - Did you know who to contact? Did you feel you could contact EFT with queries?

    If yes:

    a. What did you discuss? Did you ask for help with anything?
    b. How useful was the support received?

Overall support

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

We will contact teachers 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

Control Tutor Interview Topic Guide: Multiply Education Research Trial

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

I work for IES, an independent, not-for-profit research organisation / King's College London. Along with Ipsos and King's College London/Institute for Employment Studies, the Department for Education (DfE) has appointed our research team to understand what improves adult numeracy.

I work for IES, an independent, not-for-profit research organisation / King's College London, a public research university and higher education institution Along with Ipsos and King's College London / Institute for Employment Studies, the Department for Education has appointed our research team to evaluate the Contextualised maths/Adapted 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 Contextualised maths/Maths 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 delivering the functional skills level 1 maths qualification, learners' experiences of the course and your experiences 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?

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 describe the community that your institution serves?
3. Can you describe the typical backgrounds of your adult learners?
4. *How long have you taught the level 1 functional skills maths qualification (FSQ)?
5. *What is the duration of the FSQ course you teach? Probe 15, 17, 30 weeks or other
6. *Do you teach any other types of maths courses? Probe for age of learners
7. How many years have you been teaching?

Section B: Delivery

12. Have you heard about the **Contextualised Maths** or Realistic 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, student-driven models, and tutor facilitating student progress – i.e., key aspects of the Contextualised maths.

13. 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.

8. Tell me a bit about how you deliver functional skills.
9. What is important to keep learners engaged with maths?
10. What do adult learners struggle with within the functional skills curriculum?

a) To what extent do you use any real-world contexts to teach FSQ maths? If so, how do you use them in your teaching?

b) To what extent do you emphasise building deeper understanding of mathematical concepts and structures, and how they are interconnected when you teach FSQ maths?

11. What motivates adult learners to take or re-take the FSQ level 1 course? What do you think learners hope to gain from this?

12. What worked well / less well about delivering the FSQ course? What could be improved?

Section C: Outcomes

13. How engaged are learners with the FSQ lessons?

Probe for:

a. Attendance
b. Willingness to participate in class
c. Attitudes towards the FSQ course
d. Willingness to attempt maths problems

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

Probe for:

a. Understanding of maths concepts
b. Confidence in maths
d. Motivation for maths
e. Problem-solving skills
f. Deeper understanding and retention of knowledge
g. Improved skills for employment and life tasks

15. *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).

    a. more suited to? Why do you think that?

    b. less suited to? Why do you think that?

Section D: Enablers and barriers to participation

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

Probe for:

a. Course content

b. Staffing

c. Student attendance

d. Student barriers – e.g. SEND, ESOL learners (speech and language barriers)

e. Student attitudes or confidence

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

17. *How and when did you receive information about this trial? Can you tell us more about that communication?

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

19. *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?

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

21. Were there any challenges or barriers to being part of this trial? How do you think these barriers could be overcome in the future?

22. *What are your thoughts on participating in future research or trials like this? Would you want to be involved? Why / why not?

23. *Do you have any thoughts / recommendations as to what could be done to encourage other adult education tutors and the broader sector to take part in future trials like this one? What approaches do you think would work well?

Anything else to add?

Thank and close

Multiply FSQ Trial: Endline Learner Interview Topic Guide (Treatment)

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

I work for the Institute for Employment Studies (IES), an independent, not-for-profit research organisation / King's College London, a public research university and higher education institution. Along with Ipsos and King's College London / IES, the Department for

Education has appointed our research team to evaluate the Contextualised Maths approach for teaching level 1 maths functional skills qualification.

Today's discussion will cover your motivations for studying FSQ 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 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.

Background

1. What was your experience of learning maths prior to taking the recent Level 1 functional skills maths course?
    a. 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?
    a. [probe on delivery mode, quality of teaching, volume of content, how did it compare to original expectations]
    b. Which parts of the course, if any, did you like the most? Why?
    c. What parts of the course, if any, did you dislike the most? Why?
5. How, if at all, did this course differ to other maths courses you've taken before?
6. To what extent did your tutor help you understand the course content? How?
7. What kind of approaches did your tutor use to help you understand the course content?
    a. As relevant from response, probe key elements of both approaches to check for spillovers.
    - *Contextualised Approach:*

i. Using real life examples to help explain maths problems
ii. Encouraging you to use problem solving strategies to think through and respond to maths problems
iii. Getting students to think through maths problems together
iv. Encouraging you to listen to how other students solve problems
v. Using pictures and physical materials to make maths easier to understand

-

8. Which, if any, of the approaches helped you to understand the maths content? Why?
9. Did any of the approaches not work as well for you? Which ones and why?
10. To what extent did your tutor use real life examples?
    a. How familiar were the real-life examples that your tutor used in the sessions?
    b. In what ways, if any, did the real world examples help how you engaged with those lessons?
11. Did you attend all your course sessions?
    a. Probe: for barriers/facilitators to attending sessions
    a. Probe: were certain sessions more/less engaging? Check for Contextualised Approaches
12. Did you complete all the coursework? Probe for barriers – related to broader factors or course itself?
    a. Probe: were there certain activities that were more/less engaging? Check for Contextualised Approaches

Outcomes

13. What, if anything, did you gain from your level 1 maths course?
    a. Probe for:
        i. Improved maths outcomes
        ii. Improved understanding of maths language
        iii. Understanding of maths structures
        iv. Better understanding of connections between maths topics
        v. Confidence
        vi. Increased motivation
        vii. Willingness to attempt maths problems
        viii. Deeper understanding and retention of knowledge
        ix. Improved skills for employment and life tasks
    b. For each outcome referenced, probe on which elements of the course contributed / what their view is based on / whether they have any examples.
14. What, if anything, did you learn in the course that you didn't expect?
15. What, if any, negative impacts or challenges did you experience during the course??
16. How, if at all, has participating in this course affected your interest in continuing with further maths courses?

Enablers and barriers to participation

17. What challenges, if any, did you experience in accessing or completing your course?
18. In what ways (if any) do you think the course could be improved?
19. Trial participation: you might recall completing an online survey asking about your experiences of your course and a maths assessment, which were both part of our evaluation. Do you remember these, and how did you feel about taking part?

Closing

Anything else to add?

Confirm email details for receiving the gift voucher. Check if this can be sent to the email address they have on file.

Thank you, and close

Endline: Learner Interview Topic Guide – Control (FSQ combined trial)

Thank you for agreeing to participate in an interview about your recent experience of the functional skills level 1 maths qualification.

I work for IES, an independent, not-for-profit research organisation / King's College London, a public research university and higher education institution Along with Ipsos and King's College London / Institute for Employment Studies, the Department for Education has appointed our research team to evaluate the Contextualised maths/Adapted 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 Contextualised maths/Maths 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 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 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.

Opening

1. What was your experience with learning maths prior to taking the recent Level 1 functional skills maths course?

Experience of course

2. What was your overall impression of the course?
3. What elements of the course did you like the most? Why?
4. What elements of the course did you dislike the most? Why?
5. What kind of approaches did your tutor use to help you understand the course content?
    a. To check whether contextualised/mastery approaches are being used in BAU, probe (as relevant) for:
        i. Using real life examples to help explain maths problems
        ii. Encouraging you to use problem solving strategies to think through and respond to maths problems
        iii. Getting students to think through maths problems together
        iv. Encouraging you to listen to how other students solve problems
        v. Using pictures and physical materials to make maths easier to understand
        vi. Giving you enough time to fully understand a topic before moving on
        vii. Breaking ideas in small steps to show how they fit together and build into the larger concept and across the curriculum

viii. Showing the same idea in different ways to help you understand underlying maths principles

ix. Valuing your existing maths knowledge and helping you build on it

x. Helping you see how different maths topics connect

xi. Spending time understanding patterns and relationships between maths concepts

xii. Encouraging you to use clear mathematical language to explain your thinking

xiii. Developing a culture in which everyone believes they can succeed

6. Which of the approaches did you find worked the best for you? Why?
7. Did any of the approaches not work as well for you? Which ones and why?
8. Are there any approaches that your tutor didn't use that you think would have been helpful?
    a. Probe for Contextualised/Mastery approaches if participant raises similar concepts.

Outcomes

9. What do you think you gained from your level 1 maths course?
    a. Probe for:
        i. Improved maths outcomes
        ii. Improved understanding of maths language
        iii. Understanding of maths structures
        iv. Better understanding of connections between maths topics
        v. Confidence
        vi. Increased motivation
        vii. Willingness to attempt maths problems
        viii. Deeper understanding and retention of knowledge
        ix. Improved skills for employment and life tasks

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

10. How has participating in this course affected your interest in continuing with further maths courses?

Enablers and barriers to participation

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

12. What factors, if any, influenced how easily you were able to access or complete your course?

13. In what ways do you think the course could be improved to better support learners?

Closing

Anything else to add?

Confirm details for receiving the incentive / gift voucher. Ask for participants email address for the voucher.

Thank you and close

Multiply: FSQ Trial Stakeholder Interview Topic Guide (ETF)

Thank you for agreeing to participate in an interview about the Multiply Education Research Trials.

I work for the Institute for Employment Studies (IES), an independent, not-for-profit research organisation / King's College London, a public research university and higher education institution. Along with Ipsos and King's College London / IES, the Department for Education has appointed our research team to understand what works to improve adult numeracy.

Today's discussion will cover development and delivery of the resources, training, and support for the FSQ trial, delivery of support to providers, and your experience of being involved in the trial. 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 but may be identifiable due to the specialist role of ETF in the trial. 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.

Background, trial involvement and design

14. **Tell me briefly about your role in relation to delivery of the Multiply FSQ trial?**
    a. Probe for their involvement in:
        i. Design of learning resources
        ii. Design of training, lesson study sessions, and additional support for tutors
        iii. Contextualised, Mastery or both
15. **Can you tell me about the process of designing the training, lesson plans and materials for the contextualised arm?**
    a. What informed the design? Existing evidence, existing programmes?
    b. What worked well / less well? Are there any aspects that you think could have been improved?
16. **How did you ensure the training, lesson plans, and materials were 'trial ready', that is, sufficiently high quality to be rolled out?**
    a. Probe for:
        i. Key considerations?
        ii. Whether/how they involved tutors and learners in the process, e.g. piloting materials?
        iii. If so: what feedback did they receive? Any adjustments based on feedback ?
17. **What feedback, if any, did you receive from tutors and providers about the trial readiness packs?**
    a. Did the packs provide all the information providers and tutors needed?
    b. Were there any gaps from their perspective? What could have been added?

Delivery

Delivery of training – emphasise this is the 10 hours training in the beginning

18. **Did you deliver or observe tutor training for the FSQ trial? If so, what training did you deliver / observe (Contextualised)?**
19. **(if yes) Can you, please tell me about your experience of delivering the training sessions?**
    - Probe for:
        - Tutor understanding of materials/concepts for both Contextualised Level of consistency in delivery
        - Differences between types of providers
        - Attendance – did people attend the 4 sessions? (10 hours), barriers?
        - Engagement
        - Format (online/in person)
20. **(if yes) Were there any challenges to delivering the training? If so, what?**
21. **(if yes) What factors, if any, made it easier for you to deliver the training?**
22. **What feedback, if any, did you receive from providers/tutors about the training?**

a. Were there elements of the course that proved to be particularly challenging for tutors to grasp?
b. Were any adjustments made to the training based on the feedback?

Delivery of lesson study sessions

23. **Did you deliver or observe lesson study sessions? If so, what study sessions did you deliver / observe (Contextualised)?**
24. **(if yes) How many lesson study sessions did you deliver? To what type of tutors / providers?**
25. **(if yes) Can you please tell me about your experience delivering the lesson study sessions to tutors?**
    a. Probe:
        i. Tutor understanding of CC materials/concept
        ii. level of consistency in delivery
        iii. Differences in understanding between different types of providers
26. **(if yes) In what ways (if at all) did you adjust the delivery of the lesson study sessions as the trial went on?**
27. **(if yes) What were the challenges, if any, to delivering the lesson study sessions?**
28. **(if yes) What factors, if any, made it easier for you to deliver the lesson study sessions?**
29. **What feedback, if any, did you receive from providers about the sessions? (how, if at all, was this incorporated?)**

Additional support to tutors

30. **Can you please tell me about any other support you provided to tutors outside of the training and lesson study sessions?**
    a. Probe for: phone/email, regular/ad hoc
31. **What were the challenges, if any, to providing this support to tutors?**
32. **What factors, if any, made it easier for you to provide this support to tutors?**
33. **What feedback, if any, did you receive from tutors about this support?**

Perspectives on the views of tutors

34. **Overall, from your perspective, how engaged were tutors/providers in the trial? What makes you say that?**

•

35. **How did tutors/providers initially find the Contextualised Approach?**
    a. Similar or different to previous practices?
    b. Delivering as intended?

Views on the trial

36. **To what extent would you want to be involved in future research or trials like this? Why do you say that?**
37. **How do you think we could encourage the adult education sector to participate in more trials like this in the future? What measures could be taken?**
38. **Do you think that Contextualised Approaches could be rolled out for FSQ courses on a larger scale? Why/why not?**

b. Probe for delivering the approaches as intended, resourcing, necessary adjustments

27. What best practice would you recommend for future trials based on your experiences?

Closing

39. Do you have anything else to add or any questions for me?

Thank you and close

Multiply FSQ Trial: Stakeholder Interview Topic Guide (Etio)

Thank you for agreeing to participate in an interview about the Multiply Research Trials.

I work for the Institute for Employment Studies (IES), an independent, not-for-profit research organisation / King's College London, and Ipsos and King's College London/ IES. Along with Ipsos and King's College London / IES, the Department for Education has appointed our research team to understand what works to improve adult numeracy.

Today's discussion will last around 45 minutes and will focus on your experience of the FSQ Trial. It will cover marketing and promotion strategies, recruitment of providers, challenges faced and approaches to overcoming these.

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 but may be identifiable due to the specialist role of Etio in delivery of the FSQ trial. 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 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 the research, how your data will be used and stored, and that you are happy to participate in the interview.

Section A: Background

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

2. What experience does your organisation have working with providers in FE/ adult education sector?

Section B: Marketing and promotions

3. What marketing strategies were used to promote the FSQ trial (contextualised and Mastery arms)? (e.g. social media: X, LinkedIn, sector newsletters, publications, attendance at events, etc)

Prompts:

- Have you used different marketing strategies for each arm of the trial, if so how have these differed? Probe for contextualised and mastery arms.

4. Which strategies proved most effective in attracting providers? What worked less well? What are the learning points?

5. How did you tailor your marketing and promoting strategies to engage different types of providers?

Prompts:

- What informed your choice of marketing and promotion activities? Based on experience in the sector/best practice/piloting?
- What were some common questions during initial engagement phase; how were they addressed?
- Did your approach change during the campaign? What were the reasons for this?

6. Were any particular messages or themes resonating well / less well with potential providers? What were these messages?

Section C: Recruitment

7. What were the main challenges faced in recruiting providers to the FSQ trial?

8. How did these challenges differ across different types of learning providers?

9. How did you adapt your approach in response to these challenges?

10. What worked well/less when converting initial enquiries / EOIs into sign-ups/participation in the FSQ trial?

11. Were there any particular factors that influenced providers' decisions to participate in the trial?

Section D: Communication with providers following recruitment

12. Describe the process after providers signed up for the trial and before the trial started. What communications did providers receive, and how often?

13. How and when were the trial readiness packs distributed to the providers, tutors and learners?

14. What feedback did you receive from providers about the packs? Were there any aspects that were particularly well / less well received?

15. How often and when did you conduct calls to keep in touch with providers? What was covered during these calls? Did you contact providers outside of these calls?

Contracting and compliance

16. Can you describe the process for contracting with providers? What worked well / less well? Are there any lessons to inform contracting for this type of programme in future?

17. What strategies did you use to ensure providers adhered to the study requirements? What worked well / less well and why? How did you monitor compliance?

18. What actions are taken if a provider does not appear to comply with the requirements of the trial?

Section E: Lessons Learned

19. What lessons have you learned from working on the FSQ trial that could inform future trials?

20. How do you think we could we encourage the adult education sector to participate in more trials like this in the future? What measures could be taken?

21. What best practices would you recommend for future trials based on your experiences?

22. What would you do differently if you were to conduct a similar trial in the future?

Anything else to add?

Thank and close

Multiply FSQ Trial: Stakeholder Interview Topic Guide for Learning Providers (treatment and control)

Thank you for agreeing to participate in an interview about the Multiply Education Research Trials.

I work for the Institute for Employment Studies (IES), an independent, not-for-profit research organisation / King's College London, and Ipsos and King's College London/ IES. Along with Ipsos and King's College London / IES, the Department for Education has appointed our research team to understand what works to improve adult numeracy.

Our discussion today will cover your motivations for participating in the FSQ trial, your experiences of being part of this, including what has gone well, 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 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 the research, how your data will be used and stored, and that you are happy to participate in the interview.

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,** can you let them know that we would like to speak to them about the trial, and give them our contact details and tell them to email us?

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

Section B: Trial Communications

4. How and when did you hear about the trial?
    a. Why did your institution decide to participate?
    b. Did you consult with the teaching staff when you were making the decision to be involved? Why/Why not?

Can you talk be through the communication you received after signing up to the trial, and how it worked from your perspective? Potential prompts:

- What did you receive? (trial readiness pack?)
- When and how did you receive this?
- Did the information include all necessary information, or was anything missing?

5. How did you communicate to your staff/tutors about the trial?
    a. How did you give them information?
    b. Was this helpful in addressing their information needs? Why/why not?
    c. Did they understand what training would be involved and their role in the evaluation?
    d. Did you run additional information sessions for tutors? Any other support?

6. How / in what ways has your institution disseminated information to learners about the trial? How / in what ways has your institution ensured learners have sufficient information to decide to be involved?
    a. Did you /your staff talk them through the learner information during enrolment?
    b. Did you / your staff give them copies of the learner information?

Section C: Trial Delivery

11. 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 uploading the information to the portal?

12. Have you contacted any of the evaluation teams for support with the evaluation activities (Ipsos, IES, KCL)?

a. If yes, what did you discuss? What was helpful / less helpful about this discussion?

b. If no, were you aware that you could contact the evaluation teams for support?

13. Have you or your staff participated in Etio's regular keeping-in-touch calls?

    a. If yes, how did you find these trial support calls? Helpful/less helpful?

    b. What was discussed?

    c. If not, were you aware that Etio was delivering keeping-in-touch calls?

14. Did you receive a compliance visit from Etio?

    a. If yes, how did that go? Probe: how were you made aware of the visit, how did they prepare, access to Etio materials

    b. What was helpful / less helpful about the visit?

Section C: Training and Resources (treatment only)

15. 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?)

16. Did you feel the time for training and lesson study sessions were realistic and feasible?

17. What feedback did you receive from teaching staff regarding training expectations for the study? i.e., attendance at 3 training sessions (10 hours) and 80% of 12 weekly study sessions. Why did they say that?

18. Did you attend any of the training sessions? What did you think of the training?

If 'yes' – they also attended the training

19. Did the training give you a good understanding of the Mastery/Contextualised maths approaches? Why do you think that?

20. Did you attend any lesson study sessions?

    a. 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?

    b. Is there anything that could be changed or improved about these sessions?

Ask ALL

21. How did your institution decide which tutors would deliver Contextualised/Mastery approach following the training?

If staff had questions about delivering Contextualised/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?

   a. What did they discuss?

   b. Did they know who to contact?

   c. Did they feel they could contact ETF with queries?

22. How have staff received the Contextualised/Maths Mastery approach for FSQ level 1? Have they been interested, engaged or motivated by approach? Why do you say that?

Section D: Outcomes (treatment only)

23. In what ways, if at all, do you think the Contextualised/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?

24. What signs, if any, have you noticed that suggest a Contextualised/Mastery approach is becoming more common within your organisation?

25. Have any unintended or negative effects resulted from participating in the Contextualised/Mastery approach for teaching FSQ level 1 maths?

Section E: Views of the trial

26. What resources were needed to participate in the trial? Probe for staff time and any other costs and resources:

   a. Facilitation of the trial

   b. Delivery of the programme itself (treatment only)

27. Based on your experiences, has the programme been worthwhile? Can you share any examples? (treatment only)

28. Were there any challenges or barriers in being part of this programme and trial? If so, what were these? How could these be overcome in the future?

29. Is there anything you would change to improve the programme and trial? If so, what would that be?

30. Would you recommend that the Contextualised/Mastery approach be rolled out to more adult learning providers? Why/Why not? (treatment only)

31. What are your thoughts on participating in future research or trials like this? Why would you want to be involved, or why not?

32. How could we encourage adult education providers/the broader sector to participate in more trials or RCTs like this? What approaches do you think would work well?

Anything else to add?

Thank and close

Appendix 5: Code (randomisation and analysis)

Randomisation takes into account randomising through all relevant gateways. The steps are as follows:

1. For a new batch, combine it with old batch (for which we have retained original assignment) and randomise all units to treatment/control; randomising all units will ensure a stable assignment for the new batch units.
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 "`=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'
        }
        }
```

ANALYSIS CODE

```
***Do file for submission to DfE
**using dataset: master_file.dta

***cleaning and labelling covariates
**sex
encode ilr_sex, gen (female_yn)
la define sex_label 1 "female" 0 "male"
la values female_yn sex_label
codebook female_yn
ta female_yn, mis // 14 missing values
**ethnicity
* Combined ethnicity data ILR and survey information
*-----------------------------------------------------
gen ethnicity_n=ilr_ethnicity
/*gen, label ethnicity*/
#delimit;
la define ethnicity_n
31 "English/Welsh/Scottish/Northern Irish/British"
32 "Irish"
```

```
33 "Gypsy or Irish Traveller"
34 "Any Other White background"
35 "White and Black Caribbean"
36 "White and Black African"
37 "White and Asian"
38 "Any Other Mixed/multiple ethnic background"
39 "Indian"
40 "Pakistani"
41 "Bangladeshi"
42 "Chinese"
43 "Any other Asian background"
44 "African"
45 "Caribbean"
46 "Any other Black/African/Caribbean background"
47 "Arab"
98 "Any other ethnic group"
99 "Not provided";
#delimit cr
la val ethnicity_n ethnicity_n //assign ilr ethinicity labels
replace ethnicity_n=31 if ethnicity_n==. & svy_bs_qethnicity== "English / Welsh / Scottish / Northern Irish /
British"
replace ethnicity_n=98 if ethnicity_n==. & svy_bs_qethnicity== "Any other ethnic group, please write in"
la variable ethnicity_n "Ethnicity Combined ilr survey"
* Asian or Asian British ethnic group: Census 2021 5 categories
gen asian_asianb_n=0
replace asian_asianb_n =1 if ethnicity_n==39 //"Indian"
replace asian_asianb_n =1 if ethnicity_n==40 //"Pakistani"
replace asian_asianb_n =1 if ethnicity_n==41 //"Bangladeshi"
replace asian_asianb_n =1 if ethnicity_n==43 //"Any other Asian background"
replace asian_asianb_n =.n if ethnicity_n==99
replace asian_asianb_n =. if ethnicity_n==.
* gen Black, Black British, Caribbean or African ethnic group; Census 2021 5 categories
gen black_blackb_afri_carib_n=0
replace black_blackb_afri_carib_n =1 if ethnicity_n==44 //"African"
replace black_blackb_afri_carib_n =1 if ethnicity_n==45 //"Caribbean"
replace black_blackb_afri_carib_n =1 if ethnicity_n==46 //"Any other Black/African/Caribbean background"
replace black_blackb_afri_carib_n =.n if ethnicity_n==99
replace black_blackb_afri_carib_n =. if ethnicity_n==.
* gen Mixed or multiple ethnic groups: Census 2021 5 categories
gen mixed_multiple_n =0
replace mixed_multiple_n =1 if ethnicity_n==35 //"White and Black Caribbean"
replace mixed_multiple_n =1 if ethnicity_n==36 //"White and Black African"
replace mixed_multiple_n =1 if ethnicity_n==37 //"White and Asian"
replace mixed_multiple_n =1 if ethnicity_n==38 //"Any Other Mixed/multiple ethnic background"
replace mixed_multiple_n =.n if ethnicity_n==99
replace mixed_multiple_n =. if ethnicity_n==.
* gen white ethnic group: Census 2021 5 categories
gen white_ethnic_n=0
replace white_ethnic_n=1 if ethnicity_n==31 //"English/Welsh/Scottish/Northern Irish/British"
replace white_ethnic_n=1 if ethnicity_n==32 //"Irish"
replace white_ethnic_n=1 if ethnicity_n==33 //"Gypsy or Irish Traveller"
replace white_ethnic_n=1 if ethnicity_n==34 //"Any Other White background"
```

```
replace white_ethnic_n =.n if ethnicity_n==99
replace white_ethnic_n=. if ethnicity_n==.
* Other ethnic group: Census 2021 5 categories
gen other_ethnic_n =0
replace other_ethnic_n =1 if ethnicity_n==98 //"Any other ethnic group"
replace other_ethnic_n =1 if ethnicity_n==47 //"Arab"
replace other_ethnic_n =.n if ethnicity_n==99
replace other_ethnic_n =. if ethnicity_n==.
* gen one ethnicity var with 5 categories: Census 2021 5 categories
gen ethnicity5_n=.
replace ethnicity5_n=1 if asian_asianb_n==1 //Asian or Asian British ethnic group
replace ethnicity5_n=2 if black_blackb_afri_carib_n==1 //Black, Black British, Caribbean or African ethnic
group
replace ethnicity5_n=3 if mixed_multiple_n==1 //Mixed or multiple ethnic groups
replace ethnicity5_n=4 if white_ethnic_n==1 //White ethnic group
replace ethnicity5_n=5 if other_ethnic_n==1 //Other ethnic group
        la define ethnicity5_n 1 "Asian or Asian British ethnic group", add
        la define ethnicity5_n 2 "Black, Black British, Caribbean or African ethnic group", add
        la define ethnicity5_n 3 "Mixed or multiple ethnic groups", add
        la define ethnicity5_n 4 "White ethnic group", add
        la define ethnicity5_n 5 "Other ethnic group", add
        la var ethnicity5_n "Census 2021 ethnic groups"
        la val ethnicity5_n ethnicity5_n
*-------------------------------------
la define ethnicity_label 1 "Asian or Asian British" 2 "Black, Black British, Carribean or African" 3 "Mixed or
Multiple Ethnic Groups" 4 "White" 5 "Other Ethnic Groups"
la values ethnicity5_n ethnicity_label
ta ethnicity5_n, miss // 28 missing values
**creating binary ethnicity category
gen ethnicity_white = 1 if ethnicity5_n == 4
replace ethnicity_white = 0 if ethnicity_white == . & ethnicity5_n != .
la define ethnicity_white 1 "White" 0 "All other ethnic cat"
la values ethnicity_white ethnicity_white
ta ethnicity_white, mis //28 missing values
**age
*---------------------------------
*destring and format date of birth
*---------------------------------
destring dateofbirth, generate(dob_num)
gen dob_format = string(mod(dob_num,100), "%02.0f") + "/" + ///
        string(mod(int(dob_num/100),100), "%02.0f") + "/" + ///
        string(int(dob_num/10000), "%04.0f")
gen dob = date(dob_format, "DMY")
format dob %tdDD/NN/CCYY
* drop intermediate vars
drop dob_format dob_num dateofbirth //remove DOB identifiers
*estimate age to closest integer
gen age_date = date("01/09/2025", "DMY")
format age_date %tdDD/NN/CCYY
gen age_ies= round((age_date-dob)/365)
*---------------------------------------
**Ill-health
```

```
**insert code here for how illhealth variable was combined from ILR and survey (NICK)
*-------------------------------------
* gen var for healthproblems: ILR data
*-------------------------------------
recode ilr_llddhealthprob (2=0)(9=.n), gen (healthproblems_n)
la var healthproblems_n "Learner reported having health problems"
la copy ilr_llddhealthprob healthproblems_n
la define healthproblems_n -1 "", modify
la define healthproblems_n 9 "", modify
la define healthproblems_n 0 "Learner does not consider himself or herself to have a learning difficulty
and/or disability and/or health problem", add
la define healthproblems_n 2 "", modify
la val healthproblems_n healthproblems_n
*--------------------------------
replace healthproblems_n = . if healthproblems_n == .n
**employment status
**how was this variable derived?
replace employed_n = . if employed_n == .m
replace employed_n = . if employed_n == .n
**Region
**Changing location of the WEA to London since that was the original at randomisation
replace region = "London" if strpos(providername, "wea")
**Outcome variable construction**************************
unab aims : ilr_learnaimreftitle_aim*
gen aims=.
gen outcome=.
gen outgrade=""
forvalues i=1/20 {
replace ilr_learnaimreftitle_aim`i'=lower(ilr_learnaimreftitle_aim`i')
replace aims=`i' if strpos(ilr_learnaimreftitle_aim`i',"functional skills qualification in math") &
!strpos(ilr_learnaimreftitle_aim`i',"level 2") & !strpos(ilr_learnaimreftitle_aim`i',"3")
replace outgrade= ilr_outgrade_aim`i' if strpos(ilr_learnaimreftitle_aim`i',"functional skills qualification in
math") & ilr_outgrade_aim`i'!="" & !strpos(ilr_learnaimreftitle_aim`i',"level 2") &
!strpos(ilr_learnaimreftitle_aim`i',"3")
replace outcome= ilr_outcome_aim`i' if strpos(ilr_learnaimreftitle_aim`i',"functional skills qualification in
math") & ilr_outgrade_aim`i'!="" & !strpos(ilr_learnaimreftitle_aim`i',"level 2") &
!strpos(ilr_learnaimreftitle_aim`i',"3")
}
gen prout = .
replace prout=1 if aims!=. & outgrade=="PA"
replace prout=0 if aims!=. & outcome==3
replace prout=1 if aims!=. & outcome==1
gen prout_miss_impute = prout
replace prout_miss_impute = 0 if aims!=. & prout_miss_impute == . //Impute missing as fails
replace prout_miss_impute=. if aims!=. & outcome== 8 //Impute Code 8: Learning activities are complete
but the outcome is not yet known
gen prout_missaim_impute = prout_miss_impute
replace prout_missaim_impute = 0 if prout_missaim_impute == .
la define pass_label 1 "pass" 0 "fail", replace
la values prout pass_label
la values prout_miss_impute pass_label
la values prout_missaim_impute pass_label
```

```
**Flag non-missing aims
gen aims_nonmiss=1 if aims!=.
replace aims_nonmiss = 0 if aims_nonmiss == .
**derive course completion status
gen compstatus = .
forvalues i=1/20 {
replace compstatus = ilr_compstatus_aim`i' if strpos(ilr_learnaimreftitle_aim`i',"functional skills qualification
in math") & !strpos(ilr_learnaimreftitle_aim`i',"level 2") & !strpos(ilr_learnaimreftitle_aim`i',"3")
}
la define comp_label 1 "continuing or intending to continue learning activities leading to learning aim" 2
"completed learning activities leading to learning aim" 3 "withdrawn from learning activities leading to
learning aim" 6 "temporarily withdrawn from aim due to agreed break in learning", replace
la values compstatus comp_label
gen compstatus_fsq = compstatus
replace compstatus_fsq = 3 if compstatus_fsq == 1
replace compstatus_fsq = 1 if compstatus_fsq == 2
replace compstatus_fsq = 2 if compstatus_fsq == 6
replace compstatus_fsq = 2 if compstatus_fsq == 3
replace compstatus_fsq = 0 if compstatus_fsq == 2
la define compstatus_label 1 "completed" 0 "incomplete/withdrawn", replace
la values compstatus_fsq compstatus_label
**constructing secondary variable mean confidence
**q1 - checking change
encode svy_end_fsq_qnumconf_aq2, gen (end_conf_q1_checkchange)
replace end_conf_q1_checkchange = . if end_conf_q1_checkchange == 4
replace end_conf_q1_checkchange = 10 if end_conf_q1_checkchange == 1
replace end_conf_q1_checkchange = 20 if end_conf_q1_checkchange == 2
replace end_conf_q1_checkchange = 30 if end_conf_q1_checkchange == 3
replace end_conf_q1_checkchange = 50 if end_conf_q1_checkchange == 5
replace end_conf_q1_checkchange = 60 if end_conf_q1_checkchange == 6
replace end_conf_q1_checkchange = 1 if end_conf_q1_checkchange == 30
replace end_conf_q1_checkchange = 2 if end_conf_q1_checkchange == 50
replace end_conf_q1_checkchange = 3 if end_conf_q1_checkchange == 20
replace end_conf_q1_checkchange = 4 if end_conf_q1_checkchange == 10
replace end_conf_q1_checkchange = 5 if end_conf_q1_checkchange == 60
la define confidence_label 1 "Not at all confident" 2 "Not very confident" 3 "Neither confident nor not
confident" 4 "Fairly confident" 5 "Very confident", replace
la values end_conf_q1_checkchange confidence_label
codebook end_conf_q1_checkchange
encode svy_bs_qnumconf_aq2, gen (bs_conf_q1_checkchange)
replace bs_conf_q1_checkchange = 10 if bs_conf_q1_checkchange == 1
replace bs_conf_q1_checkchange = 20 if bs_conf_q1_checkchange == 2
replace bs_conf_q1_checkchange = 30 if bs_conf_q1_checkchange == 3
replace bs_conf_q1_checkchange = 40 if bs_conf_q1_checkchange == 4
replace bs_conf_q1_checkchange = 50 if bs_conf_q1_checkchange == 5
replace bs_conf_q1_checkchange = 60 if bs_conf_q1_checkchange == 6
replace bs_conf_q1_checkchange = 1 if bs_conf_q1_checkchange == 30
replace bs_conf_q1_checkchange = 2 if bs_conf_q1_checkchange == 50
replace bs_conf_q1_checkchange = 3 if bs_conf_q1_checkchange == 20
replace bs_conf_q1_checkchange = 4 if bs_conf_q1_checkchange == 10
replace bs_conf_q1_checkchange = 5 if bs_conf_q1_checkchange == 60
replace bs_conf_q1_checkchange = . if bs_conf_q1_checkchange == 40
```

```
la values bs_conf_q1_checkchange confidence_label
**best deals
la values svy_end_fsq_qnumconf_bq2_n confidence_label
replace svy_end_fsq_qnumconf_bq2_n = . if svy_end_fsq_qnumconf_bq2_n == .n
ren svy_end_fsq_qnumconf_bq2_n end_conf_q2_deal

encode svy_bs_qnumconf_bq2, gen(bs_conf_q2_deal)
replace bs_conf_q2_deal = 10 if bs_conf_q2_deal == 1
replace bs_conf_q2_deal = 20 if bs_conf_q2_deal == 2
replace bs_conf_q2_deal = 30 if bs_conf_q2_deal == 3
replace bs_conf_q2_deal = 40 if bs_conf_q2_deal == 4
replace bs_conf_q2_deal = 50 if bs_conf_q2_deal == 5
replace bs_conf_q2_deal = 60 if bs_conf_q2_deal == 6
replace bs_conf_q2_deal = 1 if bs_conf_q2_deal == 30
replace bs_conf_q2_deal = 2 if bs_conf_q2_deal == 50
replace bs_conf_q2_deal = 3 if bs_conf_q2_deal == 20
replace bs_conf_q2_deal = 4 if bs_conf_q2_deal == 10
replace bs_conf_q2_deal = 5 if bs_conf_q2_deal == 60
replace bs_conf_q2_deal = . if bs_conf_q2_deal == 40
***helping child with homework
encode svy_bs_qnumconf_cq2 , gen(bs_conf_q3_helpchild)
codebook bs_conf_q3_helpchildreplace bs_conf_q3_helpchild = 10 if bs_conf_q3_helpchild == 1
replace bs_conf_q3_helpchild = 20 if bs_conf_q3_helpchild == 2
replace bs_conf_q3_helpchild = 30 if bs_conf_q3_helpchild == 3
replace bs_conf_q3_helpchild = 40 if bs_conf_q3_helpchild == 4
replace bs_conf_q3_helpchild = 50 if bs_conf_q3_helpchild == 5
replace bs_conf_q3_helpchild = 60 if bs_conf_q3_helpchild == 6
replace bs_conf_q3_helpchild = 1 if bs_conf_q3_helpchild == 30
replace bs_conf_q3_helpchild = 2 if bs_conf_q3_helpchild == 50
replace bs_conf_q3_helpchild = 3 if bs_conf_q3_helpchild == 20
replace bs_conf_q3_helpchild = 4 if bs_conf_q3_helpchild == 10
replace bs_conf_q3_helpchild = 5 if bs_conf_q3_helpchild == 60
replace bs_conf_q3_helpchild = . if bs_conf_q3_helpchild == 40
la values bs_conf_q3_helpchild confidence_label
la values svy_end_fsq_qnumconf_cq2_n confidence_label
ren svy_end_fsq_qnumconf_cq2_n end_conf_q3_helpchild
***understanding interest rates
encode svy_bs_qnumconf_dq2 , gen(bs_conf_q4_interest)
replace bs_conf_q4_interest = 10 if bs_conf_q4_interest == 1
replace bs_conf_q4_interest = 20 if bs_conf_q4_interest == 2
replace bs_conf_q4_interest = 30 if bs_conf_q4_interest == 3
replace bs_conf_q4_interest = 40 if bs_conf_q4_interest == 4
replace bs_conf_q4_interest = 50 if bs_conf_q4_interest == 5
replace bs_conf_q4_interest = 60 if bs_conf_q4_interest == 6
replace bs_conf_q4_interest = 1 if bs_conf_q4_interest == 30
replace bs_conf_q4_interest = 2 if bs_conf_q4_interest == 50
replace bs_conf_q4_interest = 3 if bs_conf_q4_interest == 20
replace bs_conf_q4_interest = 4 if bs_conf_q4_interest == 10
replace bs_conf_q4_interest = 5 if bs_conf_q4_interest == 60
replace bs_conf_q4_interest = . if bs_conf_q4_interest == 40
la values bs_conf_q4_interest confidence_label
```

```
la values svy_end_fsq_qnumconf_dq2_n confidence_label
replace svy_end_fsq_qnumconf_dq2_n = . if svy_end_fsq_qnumconf_dq2_n == .n
ren svy_end_fsq_qnumconf_dq2_n end_conf_q4_interest
**keeping track of bank account balance
encode svy_bs_qnumconf_eq2 , gen(bs_conf_q5_bank)
replace bs_conf_q5_bank = 10 if bs_conf_q5_bank == 1
replace bs_conf_q5_bank = 20 if bs_conf_q5_bank == 2
replace bs_conf_q5_bank = 30 if bs_conf_q5_bank == 3
replace bs_conf_q5_bank = 40 if bs_conf_q5_bank == 4
replace bs_conf_q5_bank = 50 if bs_conf_q5_bank == 5
replace bs_conf_q5_bank = 1 if bs_conf_q5_bank == 30
replace bs_conf_q5_bank = 2 if bs_conf_q5_bank == 40
replace bs_conf_q5_bank = 3 if bs_conf_q5_bank == 20
replace bs_conf_q5_bank = 4 if bs_conf_q5_bank == 10
replace bs_conf_q5_bank = 5 if bs_conf_q5_bank == 50
la values bs_conf_q5_bank confidence_label
la values svy_end_fsq_qnumconf_eq2_n confidence_label
replace svy_end_fsq_qnumconf_eq2_n = . if svy_end_fsq_qnumconf_eq2_n == .n
ren svy_end_fsq_qnumconf_eq2_n end_conf_q5_bank
**working with numbers at job
encode svy_bs_qnumconf_fq2 , gen(bs_conf_q6_numatjob)
replace bs_conf_q6_numatjob = 10 if bs_conf_q6_numatjob == 1
replace bs_conf_q6_numatjob = 20 if bs_conf_q6_numatjob == 2
replace bs_conf_q6_numatjob = 30 if bs_conf_q6_numatjob == 3
replace bs_conf_q6_numatjob = 40 if bs_conf_q6_numatjob == 4
replace bs_conf_q6_numatjob = 50 if bs_conf_q6_numatjob == 5
replace bs_conf_q6_numatjob = 60 if bs_conf_q6_numatjob == 6
replace bs_conf_q6_numatjob = 1 if bs_conf_q6_numatjob == 30
replace bs_conf_q6_numatjob = 2 if bs_conf_q6_numatjob == 50
replace bs_conf_q6_numatjob = 3 if bs_conf_q6_numatjob == 20
replace bs_conf_q6_numatjob = 4 if bs_conf_q6_numatjob == 10
replace bs_conf_q6_numatjob = 5 if bs_conf_q6_numatjob == 60
replace bs_conf_q6_numatjob = . if bs_conf_q6_numatjob == 40
la values bs_conf_q6_numatjob confidence_label
la values svy_end_fsq_qnumconf_fq2_n confidence_label
replace svy_end_fsq_qnumconf_fq2_n = . if svy_end_fsq_qnumconf_fq2_n == .n
ren svy_end_fsq_qnumconf_fq2_n end_conf_q6_numatjob
***generating mean composite confidence
egen bs_composite_conf = rowmean( bs_conf_q1_checkchange bs_conf_q2_deal bs_conf_q3_helpchild
bs_conf_q4_interest bs_conf_q5_bank bs_conf_q6_numatjob)
egen end_composite_conf = rowmean( end_conf_q1_checkchange end_conf_q2_deal
end_conf_q3_helpchild end_conf_q4_interest end_conf_q5_bank end_conf_q6_numatjob )
******baseline and endline confidence by treatment group
**overall confidence
ta end_qconf_n treatment_n, col
ta bs_qconf_n treatment_n, col
**composite confidence questions
ta bs_conf_q1_checkchange treatment_n, col
ta end_conf_q1_checkchange treatment_n, col
ta bs_conf_q2_deal treatment_n, col
ta end_conf_q2_deal treatment_n, col
ta bs_conf_q3_helpchild treatment_n, col
```

```
ta end_conf_q3_helpchild treatment_n, col

ta bs_conf_q4_interest treatment_n, col
ta end_conf_q4_interest treatment_n, col
ta bs_conf_q5_bank treatment_n, col
ta end_conf_q5_bank treatment_n, col
ta bs_conf_q6_numatjob treatment_n, col
ta end_conf_q6_numatjob treatment_n, col
**************************************
//--BALANCE TEST--//whole sample
foreach v of varlist female_yn ethnicity_white age_ies_n healthproblems_n employed_n imd12_n
bs_composite_conf bs_qconf_n {
  reg `v' treatment_n i.region_n, vce(cluster providerid)
}
***Missing data/attrition analysis
ta compstatus_fsq if prout == . ,miss
**checks to see where do these 85 learners belong
ta providerid if compstatus_fsq == 1 & prout == .
ta providerid if providerid == 46
ta providerid if providerid == 61
ta providerid if providerid == 68
****PRIMARY OUTCOME ANALYSIS**************************
ta prout_missaim_impute treatment_n, col
**empty model controlling for region as strata and provider as level
melogit prout_missaim_impute treatment_n provider_prefcc i.region_n || providerid: , vce(cluster providerid)
***adding covars

melogit prout_missaim_impute treatment_n c.age_ies_n female_yn employed_n provider_prefcc i.region_n
|| providerid:, vce(cluster providerid)
melogit prout_missaim_impute treatment_n c.age_ies_n female_yn healthproblems_n ethnicity_white
employed_n provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
**export primary outcome model
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
*******************************************************************************
* Estimate model
*******************************************************************************
melogit prout_missaim_impute treatment_n c.age_ies_n female_yn healthproblems_n ethnicity_white
employed_n provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
*******************************************************************************
* Extract estimates
*******************************************************************************
scalar N_m = e(N)
scalar b_m = _b[treatment_n]
scalar se_m = _se[treatment_n]
scalar z  = b_m / se_m
scalar pval = 2*normal(-abs(z))
scalar lo_m = b_m - 1.96*se_m
scalar hi_m = b_m + 1.96*se_m
```

```
*******************************************************************************
* Marginal predicted probabilities at treatment=0 vs 1
*******************************************************************************
quietly margins, at(treatment_n=(0 1)) predict(mu)
matrix T = r(table)
scalar p0_m = T[1,1]
scalar p1_m = T[1,2]
*******************************************************************************
* Cohen's h
*******************************************************************************
scalar cohen_h = 2*asin(sqrt(p1_m)) - 2*asin(sqrt(p0_m))
scalar se_h  = sqrt(2/N_m)
scalar h_lo_m = cohen_h - 1.96*se_h
scalar h_hi_m = cohen_h + 1.96*se_h
*******************************************************************************
* Post results (Hedges g information left empty with .)
*******************************************************************************
post `M' ("melogit_prout") ("prout_missaim_impute") ("melogit") (N_m) (b_m) (se_m) (lo_m) (hi_m) (p0_m)
(p1_m) (cohen_h) (h_lo_m) (h_hi_m) (.) (.) (.) (.) (pval)
postclose `M'
preserve
use `modelres', clear
save "P:\PROJECTS\6304T2 - Functional Skills Multiply Trial\Research\Data analysis\Impact
Evaluation\primary_model_results.dta", replace
restore

***Secondary outcome analysis
replace bs_qconf_n = . if bs_qconf_n == .n
replace end_qconf_n = . if end_qconf_n == .n
ta bs_qconf_n treatment_n
ta end_qconf_n treatment_n
sum bs_qconf_n
bysort treatment_n: summ bs_qconf_n
sum end_qconf_n
bysort treatment_n: summ end_qconf_n
**composite confidence
sum bs_composite_conf
bysort treatment_n: summ bs_composite_conf
sum end_composite_conf
bysort treatment_n: summ end_composite_conf
***distribution
tempname G1 G2 G3 G4
twoway (histogram bs_composite_conf if treatment_n==0, width(1) start(1) color(navy%90) ), title("Control
at Baseline") ytitle("Density") xtitle("Confidence in Maths (composite score)") name(`G1', replace)
twoway (histogram bs_composite_conf if treatment_n==1, width(1) start(1) color(navy%90) ),
title("Treatment at Baseline") ytitle("Density") xtitle("Confidence in Maths (composite score)") name(`G2',
replace)
twoway (histogram end_composite_conf if treatment_n==0, width(1) start(1) color(navy%90) ), title("Control
at Endline") ytitle("Density") xtitle("Confidence in Maths (composite score)") name(`G3', replace)
twoway (histogram end_composite_conf if treatment_n==1, width(1) start(1) color(navy%90) ),
title("Treatment at Endline") ytitle("Density") xtitle("Confidence in Maths (composite score)") name(`G4',
replace)
```

```
graph combine `G1' `G2' `G3' `G4', col(2) imargin(tiny)
tempname G1 G2 G3 G4
twoway (histogram bs_qconf_n if treatment_n==0, width(1) start(1) color(navy%90) ), title("Control at
Baseline") ytitle("Density") xtitle("Confidence in Maths") name(`G1', replace)
twoway (histogram bs_qconf_n if treatment_n==1, width(1) start(1) color(navy%90) ), title("Treatment at
Baseline") ytitle("Density") xtitle("Confidence in Maths") name(`G2', replace)
twoway (histogram end_qconf_n if treatment_n==0, width(1) start(1) color(navy%90) ), title("Control at
Endline") ytitle("Density") xtitle("Confidence in Maths") name(`G3', replace)
twoway (histogram end_qconf_n if treatment_n==1, width(1) start(1) color(navy%90) ), title("Treatment at
Endline") ytitle("Density") xtitle("Confidence in Maths") name(`G4', replace)
graph combine `G1' `G2' `G3' `G4', col(2) imargin(tiny)
**secondary outcome regressions
mixed end_qconf_n treatment bs_qconf_n i.region_n|| providerid: , vce(cluster providerid)
mixed end_qconf_n treatment_n bs_qconf_n c.age_ies_n female_yn healthproblems_n ethnicity_white
employed_n provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
**export secondary outcome model
mixed end_qconf_n treatment_n bs_qconf_n c.age_ies_n female_yn healthproblems_n ethnicity_white
employed_n provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
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 mean0 double mean1 double md double sd_res double
hedges_g double g_lo double g_hi double pval using `modelres', replace
scalar N_m = e(N)
scalar b_m = _b[treatment_n]
scalar se_m = _se[treatment_n]
scalar lo_m = b_m - 1.96*se_m
scalar hi_m = b_m + 1.96*se_m
scalar z  = b_m / se_m
scalar pval = 2*normal(-abs(z))
quietly margins, at(treatment_n=(0 1))
matrix M = r(table)
scalar mean0 = M[1,1]
scalar mean1 = M[1,2]
scalar md  = mean1 - mean0
capture scalar sd_res = exp(_b[lnsig_e:_cons])
if _rc {
  di as err "Could not read residual SD from lnsig_e; computing from residuals as fallback"
  predict double ehat, residuals
  quietly summarize ehat
  scalar sd_res = r(sd)
  drop ehat
}
scalar J   = 1 - (3 / (4*N_m - 9))
scalar hedges_g = J * (md / sd_res)
scalar se_g   = J * (se_m / sd_res)
scalar g_lo   = hedges_g - 1.96 * se_g
scalar g_hi   = hedges_g + 1.96 * se_g
post `M' ("mixed_endconf") ("end_qconf_n") ("mixed") (N_m) (b_m) (se_m) (lo_m) (hi_m) (mean0) (mean1)
(md) (sd_res) (hedges_g) (g_lo) (g_hi) (pval)
postclose `M'
preserve
```

```
use `modelres', clear
list, noobs abbreviate(20)
save "P:\PROJECTS\6304T2 - Functional Skills Multiply Trial\Research\Data analysis\Impact
Evaluation\secondary1_model_results.dta", replace
restore
**********Secondary outcome - composite confidence
mixed end_composite_conf treatment_n bs_composite_conf i.region_n || providerid: , vce(cluster
providerid)
mixed end_composite_conf treatment_n bs_composite_conf c.age_ies_n female_yn healthproblems_n
ethnicity_white employed_n provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
**export secondary outcome model2
mixed end_composite_conf treatment_n bs_composite_conf c.age_ies_n female_yn healthproblems_n
ethnicity_white employed_n provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
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 mean0 double mean1 double md double sd_res double
hedges_g double g_lo double g_hi double pval using `modelres', replace
scalar N_m = e(N)
scalar b_m = _b[treatment_n]
scalar se_m = _se[treatment_n]
scalar lo_m = b_m - 1.96*se_m
scalar hi_m = b_m + 1.96*se_m
scalar z  = b_m / se_m
scalar pval = 2*normal(-abs(z))
quietly margins, at(treatment_n=(0 1))
matrix M = r(table)
scalar mean0 = M[1,1]
scalar mean1 = M[1,2]
scalar md  = mean1 - mean0
capture scalar sd_res = exp(_b[lnsig_e:_cons])
if _rc {
  di as err "Could not read residual SD from lnsig_e; computing from residuals as fallback"
  predict double ehat, residuals
  quietly summarize ehat
  scalar sd_res = r(sd)
  drop ehat
}
scalar J   = 1 - (3 / (4*N_m - 9))
scalar hedges_g = J * (md / sd_res)
scalar se_g   = J * (se_m / sd_res)
scalar g_lo   = hedges_g - 1.96 * se_g
scalar g_hi   = hedges_g + 1.96 * se_g
post `M' ("mixed_endconf") ("mean_end_qnumconf") ("mixed") (N_m) (b_m) (se_m) (lo_m) (hi_m) (mean0)
(mean1) (md) (sd_res) (hedges_g) (g_lo) (g_hi) (pval)
postclose `M'
preserve
use `modelres', clear
list, noobs abbreviate(20)
save "P:\PROJECTS\6304T2 - Functional Skills Multiply Trial\Research\Data analysis\Impact
Evaluation\secondary2_model_results.dta", replace
restore
```

```
***Investigating each confidence question behind mean composite confidence
mixed end_conf_q1_checkchange treatment_n bs_conf_q1_checkchange c.age_ies_n female_yn
healthproblems_n ethnicity_white employed_n provider_prefcc imd12_n i.region_n || providerid:,
vce(cluster providerid)
mixed end_conf_q2_deal treatment_n bs_conf_q2_deal c.age_ies_n female_yn healthproblems_n
ethnicity_white employed_n provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
mixed end_conf_q3_helpchild treatment_n bs_conf_q3_helpchild c.age_ies_n female_yn
healthproblems_n ethnicity_white employed_n provider_prefcc imd12_n i.region_n || providerid:,
vce(cluster providerid)
mixed end_conf_q4_interest treatment_n bs_conf_q4_interest c.age_ies_n female_yn healthproblems_n
ethnicity_white employed_n provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
mixed end_conf_q5_bank treatment_n bs_conf_q5_bank c.age_ies_n female_yn healthproblems_n
ethnicity_white employed_n provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
mixed end_conf_q6_numatjob treatment_n bs_conf_q6_numatjob c.age_ies_n female_yn
healthproblems_n ethnicity_white employed_n provider_prefcc imd12_n i.region_n || providerid:,
vce(cluster providerid)
*******Interaction effects****************************************
**Sex
melogit prout_missaim_impute treatment_n##female_yn c.age_ies_n ethnicity_white healthproblems_n
employed_n provider_prefcc imd12 i.region_n || providerid:, vce(cluster providerid)
**age
melogit prout_missaim_impute treatment_n##c.age_ies_n female_yn c.age_ies_n ethnicity_white
healthproblems_n employed_n provider_prefcc imd12 i.region_n || providerid:, vce(cluster providerid)
**health
melogit prout_missaim_impute treatment_n##healthproblems_n c.age_ies_n female_yn c.age_ies_n
ethnicity_white employed_n provider_prefcc imd12 i.region_n || providerid:, vce(cluster providerid)
**ethnicity
melogit prout_missaim_impute treatment_n##ethnicity_white healthproblems_n c.age_ies_n female_yn
c.age_ies_n employed_n provider_prefcc imd12 i.region_n || providerid:, vce(cluster providerid)
***employment
melogit prout_missaim_impute treatment_n##employed_n ethnicity_white healthproblems_n c.age_ies_n
female_yn c.age_ies_n provider_prefcc imd12 i.region_n || providerid:, vce(cluster providerid)
**deprivation
melogit prout_missaim_impute treatment_n##imd12 employed_n ethnicity_white healthproblems_n
c.age_ies_n female_yn c.age_ies_n provider_prefcc i.region_n || providerid:, vce(cluster providerid)
***CACE analysis*************************************************************
**drop empty session vars
drop session1n_c session1n_hl session1n_p session2ssm_c session2ssm_hl session2ssm_p session3s_c
session3s_hl session3s_p session4ssm2_c session4ssm2_hl session4ssm2_p session5n2_c
session5n2_hl session5n2_p session6s2_c session6s2_hl session6s2_p

local sessions session1 session2 session3 session4 session5 session6 session7 session8 session9
session10 session11 session12
foreach i of local sessions {
 replace `i' = lower(trim(`i'))
 encode `i', gen(`i'_n)
}
local sessions session1_n session2_n session3_n session4_n session5_n session6_n session7_n
session8_n session9_n session10_n session11_n session12_n
foreach i of local sessions {
 recode `i' (1=.)(2=1)(3=2)(4=3)
 label define `i' 1 "Attended all", modify
```

```
 label define `i' 2 "Attended part", modify
 label define `i' 3 "Did not attend", modify
}
**generating binary attendance var for full/part
local sessions session1_n session2_n session3_n session4_n session5_n session6_n session7_n
session8_n session9_n session10_n session11_n session12_n
foreach i of local sessions {
 gen attend_`i'= 1 if `i' == 1
 replace attend_`i'= 1 if `i' == 2
 replace attend_`i'= 0 if `i' == .
 replace attend_`i'= 0 if `i' == 3
}
la define yn_label 1 "yes" 0 "no", replace
la values attend_session1_n yn_label
la values attend_session2_n yn_label
la values attend_session3_n yn_label
la values attend_session4_n yn_label
la values attend_session5_n yn_label
la values attend_session6_n yn_label
la values attend_session7_n yn_label
la values attend_session8_n yn_label
la values attend_session9_n yn_label
la values attend_session10_n yn_label
la values attend_session11_n yn_label
la values attend_session12_n yn_label

**calculating total attendance (dosage)
gen dosage = attend_session1_n + attend_session2_n + attend_session3_n + attend_session4_n +
attend_session5_n + attend_session6_n + attend_session7_n + attend_session8_n + attend_session9_n +
attend_session10_n + attend_session11_n + attend_session12_n
gen itt_treated=0
replace itt_treated= 1 if dosage>=10
replace itt_treated = 0 if itt_treated == .
la values itt_treated yn_label
ta itt_treated if treatment_n == 1
*** CACE Contextualised FSQ1 full model with clustered errors
ivregress 2sls prout_missaim_impute c.age_ies_n female_yn imd12 employed_n healthproblems_n
ethnicity_white provider_prefcc i.region_n (itt_treated=treatment_n), vce(cluster providerid)
ivregress 2sls end_qconf_n bs_qconf_n c.age_ies_n female_yn imd12 employed_n healthproblems_n
ethnicity_white provider_prefcc i.region_n (itt_treated=treatment_n), vce(cluster providerid)
ivregress 2sls end_composite_conf bs_composite_conf c.age_ies_n female_yn imd12 employed_n
healthproblems_n ethnicity_white provider_prefcc i.region_n (itt_treated=treatment_n), vce(cluster
providerid)
***Missing Data Analysis
* Create dummies to check >5% missing
local missvars prout prout_miss_impute female_yn imd12_n employed_n healthproblems_n ethnicity5_n
end_qconf_n bs_qconf_n bs_composite_conf end_composite_conf
foreach i of local missvars {
        gen `i'_miss = 0
        replace `i'_miss = 1 if `i'==. | `i'==.n |`i'==.m
}
```

```
***Finding; missingness is <5% for all demographic vars - ethnicity, employment, sex, ill-health .
Missingness is >5% on outcome vars
* Test predictability of missing covars // does treatment predict missingness in each covar
local missdummies prout_miss female_yn_miss imd12_n_miss employed_n_miss healthproblems_n_miss
ethnicity5_n_miss
foreach i of local missdummies {
        logit `i' treatment_n, vce(cluster providerid)
}
*test predicatability of missing in primary and secondary outcome
logit prout_miss treatment_n c.age_ies_n female_yn imd12_n employed_n healthproblems_n
ethnicity_white provider_prefcc ib3.region_n i.providerid, vce(cluster providerid)
logit end_qconf_n_miss treatment_n bs_qconf_n c.age_ies_n female_yn imd12_n employed_n
healthproblems_n ethnicity_white provider_prefcc ib3.region_n i.providerid, vce(cluster providerid)
logit end_composite_conf_miss treatment_n bs_composite_conf c.age_ies_n female_yn imd12_n
employed_n healthproblems_n ethnicity_white provider_prefcc ib3.region_n i.providerid, vce(cluster
providerid)
***Imputing missing vars that are only background characteristics
misstable sum prout prout_miss_impute female_yn age_ies_n imd12 provider_prefcc employed_n
healthproblems_n ethnicity5_n region_n end_qconf_n bs_qconf_n bs_composite_conf
end_composite_conf


preserve
mi set wide
mi register imputed female_yn employed_n healthproblems_n ethnicity_white imd12_n bs_qconf_n
bs_composite_conf
mi register regular treatment_n age_ies_n region_n providerid
mi impute chained (regress) bs_qconf_n bs_composite_conf (logit, augment) female_yn employed_n
healthproblems_n ethnicity_white imd12_n=age_ies_n treatment_n i.providerid i.region_n, add(10)
rseed(123) force
mi estimate, post: mixed prout_missaim_impute treatment_n age_ies_n female_yn imd12_n employed_n
healthproblems_n ethnicity_white provider_prefcc i.region_n || providerid:, vce(cluster providerid)
mi estimate, post: mixed end_qconf_n treatment_n bs_qconf_n age_ies_n female_yn imd12_n
employed_n healthproblems_n ethnicity_white provider_prefcc i.region_n || providerid:, vce(cluster
providerid)
mi estimate, post: mixed end_composite_conf treatment_n bs_composite_conf age_ies_n female_yn
imd12_n employed_n healthproblems_n ethnicity_white provider_prefcc i.region_n || providerid:,
vce(cluster providerid)
restore
mi estimate, post: mixed prout_missaim_impute treatment_n age_ies_n female_yn imd12_n employed_n
healthproblems_n ethnicity_white provider_prefcc i.region_n || providerid:, vce(cluster providerid)
mi estimate, post: mixed end_qconf_n treatment_n bs_qconf_n age_ies_n female_yn imd12_n
employed_n healthproblems_n ethnicity_white provider_prefcc i.region_n || providerid:, vce(cluster
providerid)
mi estimate, post: mixed end_composite_conf treatment_n bs_composite_conf age_ies_n female_yn
imd12_n employed_n healthproblems_n ethnicity_white provider_prefcc i.region_n || providerid:,
vce(cluster providerid)
**Additional sensitivity analysis
**dropping 65 learners
melogit prout_miss_impute treatment_n c.age_ies_n female_yn healthproblems_n ethnicity_white
employed_n provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
```

```
gen prout_miss_impute2 = prout_miss_impute
(65 missing values generated)
replace prout_miss_impute2 = . if prout == . & compstatus_fsq == 1
(85 real changes made, 85 to missing)
ta prout_miss_impute2, miss
**further dropping 85 learners
melogit prout_miss_impute2 treatment_n c.age_ies_n female_yn healthproblems_n ethnicity_white
employed_n provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
melogit prout_miss_impute2 treatment_n c.age_ies_n female_yn healthproblems_n ethnicity_white
employed_n provider_prefcc imd12_n
**dropping all missing observations
melogit prout treatment_n c.age_ies_n female_yn healthproblems_n ethnicity_white employed_n
provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
***CACE/LATE analysis for Compstatus
melogit compstatus_fsq treatment_n c.age_ies_n female_yn healthproblems_n ethnicity_white employed_n
provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
predict double xb1, xb
gen double phat = invlogit(xb1)
gen double uhat = compstatus_fsq - phat
melogit prout_missaim_impute treatment_n compstatus_fsq uhat c.age_ies_n i.female_yn i.employed_n
healthproblems_n ethnicity_white imd12_n provider_prefcc i.region_n || providerid:, vce(cluster providerid)
***Effect size - Cohen's h
**Primary Outcome**FSQ maths attainment
melogit prout_missaim_impute treatment_n c.age_ies_n female_yn healthproblems_n ethnicity_white
employed_n provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)

scalar N_m = e(N)
scalar b_m = _b[treatment_n]
scalar se_m = _se[treatment_n]
scalar z = b_m / se_m
scalar pval = 2*normal(-abs(z))
scalar lo_m = b_m - 1.96*se_m
scalar hi_m = b_m + 1.96*se_m
margins, at(treatment_n=(0 1)) predict(mu)
matrix T = r(table)
scalar p0_m = T[1,1]
scalar p1_m = T[1,2]
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
display "------------------------------------------"
display "N: " N_m
display "Treatment coef: " b_m
display "SE: " se_m
display "95% CI: [" lo_m ", " hi_m "]"
display "p-value: " pval
display "p0 (control predicted prob): " p0_m
display "p1 (treat predicted prob): " p1_m
display "Cohen's h: " h_m
display "Cohen's h 95% CI: [" h_lo_m ", " h_hi_m "]"
display "------------------------------------------"
```

```
***Effect size confidence in maths (secondary outcome1)
mixed end_qconf_n treatment_n bs_qconf_n c.age_ies_n female_yn healthproblems_n ethnicity_white
employed_n provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
scalar N_m = e(N)
scalar b_m = _b[treatment_n]
scalar se_m = _se[treatment_n]
scalar lo_m = b_m - 1.96*se_m
scalar hi_m = b_m + 1.96*se_m
scalar z  = b_m / se_m
scalar pval = 2*normal(-abs(z))
quietly margins, at(treatment_n=(0 1))
matrix M = r(table)
scalar mean0 = M[1,1]
scalar mean1 = M[1,2]
scalar md  = mean1 - mean0
capture scalar sd_res = exp(_b[lnsig_e:_cons])
if _rc {
  di as err "Could not read residual SD from lnsig_e; computing from residuals as fallback"
  predict double ehat, residuals
  quietly summarize ehat
  scalar sd_res = r(sd)
  drop ehat
}

scalar J   = 1 - (3 / (4*N_m - 9))
scalar hedges_g = J * (md / sd_res)
scalar se_g   = J * (se_m / sd_res)
scalar g_lo   = hedges_g - 1.96 * se_g
scalar g_hi   = hedges_g + 1.96 * se_g
display "----------------------------------------------"
display "N: " N_m
display "Treatment coef: " b_m
display "SE: " se_m
display "95% CI: [" lo_m ", " hi_m "]"
display "p-value: " pval
display "Hedge's g: " hedges_g
display "95% CI: [" g_lo ", " g_hi "]"
display "----------------------------------------------"
***Effect size confidence in maths (composite)
mixed end_composite_conf treatment_n bs_composite_conf c.age_ies_n female_yn healthproblems_n
ethnicity_white employed_n provider_prefcc imd12_n i.region_n || providerid:, vce(cluster providerid)
scalar N_m = e(N)
scalar b_m = _b[treatment_n]
scalar se_m = _se[treatment_n]
scalar lo_m = b_m - 1.96*se_m
scalar hi_m = b_m + 1.96*se_m
scalar z  = b_m / se_m
scalar pval = 2*normal(-abs(z))

quietly margins, at(treatment_n=(0 1))
matrix M = r(table)
scalar mean0 = M[1,1]
```

```
scalar mean1 = M[1,2]
scalar md  = mean1 - mean0
capture scalar sd_res = exp(_b[lnsig_e:_cons])
if _rc {
  di as err "Could not read residual SD from lnsig_e; computing from residuals as fallback"
  predict double ehat, residuals
  quietly summarize ehat
  scalar sd_res = r(sd)
  drop ehat
}
scalar J   = 1 - (3 / (4*N_m - 9))
scalar hedges_g = J * (md / sd_res)
scalar se_g   = J * (se_m / sd_res)
scalar g_lo   = hedges_g - 1.96 * se_g
scalar g_hi   = hedges_g + 1.96 * se_g
display "------------------------------------------------"
display "N: " N_m
display "Treatment coef: " b_m
display "SE: " se_m
display "95% CI: [" lo_m ", " hi_m "]"
display "p-value: " pval
display "Hedge's g: " hedges_g
display "95% CI: [" g_lo ", " g_hi "]"
display "------------------------------------------------"
***calculating ICC using primary model
**This is the cluster level variance (between provider variance)
**var(_cons) = 2.393732
**In logistic mixed models, the residual variance is fixed
**Residual variance = 3.29 (fixed)
display ICC = 2.393732/ (2.393732+3.29)
**result = 0.42 = 42%
***ex-post power calculation (to conduct a full scale RCT what will be the number of providers required?)
**Assumptions
**ICC = 0.42 # intra-cluster correlation
**m = 25 # avg cluster size
**D = 1+(ICC*(m-1)) # design effect
**To adjust D for attrition use m* = m(1-Ym) //Ym is the rate of attrition at individual level, but here we do
not adjust for that
**Cohen's h = 0.2
**Cluster k (provider) attrition in pilot trial - 42%
**pooled SD = 1
**Calculate the required sample size using Campbell et al. (2000),Chapter 20 of Gelman and Hill (2006)
and Borm et al. (2007) fornula as below
**n = [(4*SD^2(Z(1-a/2) + Z(1-b))^2)/h^2]*[1+ICC(m-1)]  //where, n is the sample size; and
**OR
****k(1 − Yk) = [(4*SD^2(2.8*2.8))]*[1+ICC(m-1)]/(h^2*m)  // SD is the pooled standard deviation; a=0.05
(95% significance); b=0.8 (power); h is Cohen's h, z(0.95) = 1.96 + z(0.2) = 2.8; k is the number of total
clusters, Yk is provider attrition rate, [1+ICC(m-1)] is the design effect
***k(1 − Yk) = ((4*1*1*(2.8*2.8))*(1+(0.42*(25-1)))/(0.2*0.2*25)
****k(1 − Yk) = [(4*2.8*2.8)*(1+(0.42*24))]/(0.2*0.2*25)
display [(4*2.8*2.8)*(1+(0.42*24))]/(0.2*0.2*25)
347.4688
```

**add provider level attrition of 42% (1-Yk = 0.58)
display [(4*2.8*2.8)*(1+(0.42*24))]/(0.2*0.2*25*0.58)
599.08414
**using the actual effect size observed in pilot of 0.02 and pooled SD of 0.5
. display [(4*0.5*0.5*2.8*2.8)*(1+(0.42*24))]/(0.02*0.02*25*0.58)
14977.103

Appendix 6: Additional tables/charts discussed in the text

Table A6 - 1: Sample proportions by baseline category

Variable	Term	N	Percentage (%)
Deprivation	Missing	15	[u]
Deprivation	Low	468	65%
Deprivation	High	236	33%
Employment	Missing	27	[u]
Employment	Not Employed	345	48%
Employment	Employed	347	48%
Ethnicity	Missing	28	[u]
Ethnicity	White	384	53%
Ethnicity	Mixed/Multiple ethnic groups	29	[u]
Ethnicity	Asian or Asian British	87	12%
Ethnicity	Black, Black British, Carib-bean or African	124	17%
Ethnicity	Other ethnic group	67	9%
Sex	Missing	[c]	[c]
Sex	Male	151	21%
Sex	Female	566	79%
Ill health	Missing	14	[u]
Ill health	Unknown	7	[u]
Ill health	No	497	69%
Ill health	Yes	201	28%
Region	East Midlands	103	14%
Region	East of England	295	41%
Region	London	142	17%
Region	North East	[c]	[c]
Region	North West	82	11%
Region	South East	57	8%
Region	South West	[c]	[c]
Region	Yorkshire and the Humber	25	[u]
Maths confidence (baseline)	Missing	524	73%
Maths confidence (baseline)	(1) Not at all confident	9	[u]
Maths confidence (baseline)	(2) Not very confident	24	[u]
Maths confidence (baseline)	(3) Neither confident nor not confident	36	5%
Maths confidence (baseline)	(4) Fairly confident	83	12%
Maths confidence (baseline)	(5) Very confident	43	6%
Maths confidence (endline)	Missing	459	64%
Maths confidence (endline)	(1) Not at all confident	5	[u]
Maths confidence (endline)	(2) Not very confident	5	[u]

Variable	Term	N	Percentage (%)
Maths confidence (endline)	**(3) Neither confident nor not confident**	17	[u]
Maths confidence (endline)	**(4) Fairly confident**	136	19%
Maths confidence (endline)	**(5) Very confident**	97	13%

Note: Shorthand [u] indicates percentages have been suppressed due to low reliability of cell counts (where n < 30)

Source: ILR and survey data. Base: All learners

Table A6 - 2: Balance of baseline characteristics

Covariate	Estimate	SE	P-value	N
Female (reference: Male)	0.08	0.04	0.084	705
White (reference: others)	-0.14	0.12	0.274	691
Ill health (reference: no)	0.03	0.05	0.566	698
Most deprived (reference: no)	0.08	0.10	0.438	704
Employed (reference: no)	-0.23*	0.09	0.020	692
Age	-0.15	0.81	0.859	719
Confidence in maths (composite)	0.20	0.21	0.361	194
Confidence in maths (overall)	0.10	0.13	0.436	195

Note: Note: Significance Levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors clustered at provider level, with region fixed effects. Source: ILR and baseline survey. Base: All learners, 719

Table A6 - 3: Achievement of secondary outcomes by trial arm

Secondary outcomes	Treatment (n/N)	Treatment Mean (SD)	Control (n/N)	Control Mean (SD)	Total (n/N)	Total Mean (SD)
Confidence in maths (Baseline)	72 (35%)	3.76 (1.07)	123 (24%)	3.59 (1.12)	195 (27%)	3.65 (1.09)
Confidence in maths (Endline)	92 (45%)	4.27 (0.81)	168 (33%)	4.18 (0.80)	260 (36%)	4.21 (0.80)
Composite confidence (Baseline)	73 (36%)	4.08 (0.80)	121 (24%)	3.88 (0.86)	194 (27%)	3.96 (0.84)
Composite confidence (Endline)	92 (45%)	4.37 (0.73)	168 (33%)	4.33 (0.69)	260 (36%)	4.35 (0.71)

Note: Source: Baseline and endline survey. Base: All learners, 719

Table A6 - 4: Regression results for primary and secondary outcome analysis

Variables	Primary outcome	Confidence in maths	Composite confidence
Treatment	-0.05	-0.17	0.24
Treatment	(0.97)	(0.14)	(0.18)
Baseline confidence in maths	x	0.25***	x
Baseline confidence in maths	x	(0.04)	x
Baseline composite confidence	x	x	0.28**
Baseline composite confidence	x	x	(0.11)
Age	0.02**	0.00	0.00
Age	(0.01)	(0.00)	(0.00)
Female (reference: male)	0.07	-0.27	0.43*
Female (reference: male)	(0.30)	(0.14)	(0.18)
Ill-health (reference: no)	-0.11	-0.27	-0.28
Ill-health (reference: no)	(0.33)	(0.11)	(0.17)
White (reference: others)	0.18	0.05	0.21
White (reference: others	(0.18)	(0.09)	(0.17)
Employed (reference: no)	0.01	-0.12	0.20*
Employed (reference: no)	(0.21)	(0.10)	(0.10)
Most deprived (reference: no)	0.03	0.03	-0.35*
Most deprived (reference: no)	(0.14)	(0.17)	0.14)
CA preference (= yes)	0.49	-0.14	-0.26
CA preference (= yes)	(0.69)	(0.21)	(0.28)
Region	x	x	x
East of England	-1.87	-0.03	-0.33*
East of England	(1.13)	(0.18)	(0.13)
London	-1.37	0.10	0.09
London	(0.97)	(0.12)	(0.20)
North East	0.00	x	x
North East	x	x	x
Nort West	-0.92	0.16	0.24
Nort West	(0.92)	(0.15)	(0.48)
South East	-0.72	0.29	-0.28
South East	(1.46)	(0.25)	(0.27)
South West	-0.65	x	x
South West	(0.90)	x	x
Yorkshire and the Humber	-1.79	-0.12	-0.04
Yorkshire and the Humber	(1.35)	(0.12)	(0.13)
N	658	76	75
Constant	0.00	3.60***	2.99***
Constant	(0.59)	(0.30)	(0.76)
Provider variance (constant)	2.35	0.00	0.00
Provider variance (constant)	(1.48)	(0.00)	(0.00)

Variables	Primary outcome	Confidence in maths	Composite confidence
Residual variance	x	0.25	0.35
Residual variance	x	(0.05)	(0.10)

Note: Significance Levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors clustered at provider level, with region fixed effects. Standard errors presented in parathesis. Notation x indicates data were unavailable. Source: ILR and baseline survey. Base: All learners for primary outcome, 719. Only survey respondents for secondary outcome confidence in maths, 76 for overall confidence in maths and 75 for composite confidence in maths.

Table A6 - 5: Regression results for interaction effects

Variables	Female	Age	Ill-health	Ethnicity	Employment status	Deprivation
Treatment	-0.47	-0.36	-0.33	-0.13	0.00	0.08
Treatment	(1.15)	(1.22)	(1.00)	(0.98)	(0.98)	(1.03)
Female (reference: male)	-0.02	0.06	0.05	0.07	0.06	0.07
Female (reference: male)	(0.34)	(0.30)	(0.30)	(0.30	(0.30)	(0.30)
Age	0.02**	0.02*	0.02*	0.02***	0.02**	0.02**
Age	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
White (reference: others)	0.19	0.19	0.18	0.13	0.18	0.19
White (reference: others)	(0.17)	(0.18)	(0.16)	(0.14)	(0.18)	(0.18)
Ill-health (reference: no)	-0.12	-0.12	-0.29	-0.12	-0.11	-0.11
Ill-health (reference: no)	(0.33)	(0.34)	(0.44)	(0.33)	(0.34)	(0.33)
Employed (reference: no)	0.03	0.02	-0.02	0.01	0.04	0.01
Employed (reference: no)	(0.21)	(0.21)	(0.23)	(0.21)	(0.26)	(0.21)
Most deprived (reference: no)	0.04	0.03	0.01	0.03	0.03	0.09
Most deprived (reference: no)	(0.14)	(0.14)	(0.14)	(0.15)	(0.14)	(0.14)
London	-1.34	-1.38	-1.43	-1.38	-1.36	-1.40
London	(0.96)	(0.97)	91.00)	(0.97)	(0.97)	(0.97)
North East	x	x	x	x	x	x
North East	x	x	x	x	x	x
North West	-0.91	-0.92	-1.00	-0.93	-0.92	-0.92
North West	(0.91)	(0.93)	(0.98)	(0.93)	(0.92)	(0.92)
South East	-0.67	-0.70	-0.84	-0.74	-0.73	-0.75
South East	91.49)	91.49)	91.46)	(1.46)	(1.47)	(1.45)
South West	-0.47	-0.58	-0.97	-0.71	-0.58	-0.70
South West	(0.92)	(0.93)	(0.89)	(0.94)	(0.93)	(0.92)
Yorkshire and the Humber	-1.75	-1.79	-1.88	-1.80	-1.81	-1.76
Yorkshire and the Humber	(1.34)	(1.35)	(1.38)	(1.36)	(1.36)	(1.32)
Treatment*female	0.49	x	x	x	x	x
Treatment*female	(0.62)	x	x	x	x	x
Treatment*age	x	0.01	x	x	x	x
Treatment*age	x	(0.02)	x	x	x	x
Treatment*ill-health	x	x	0.80	x	x	x
Treatment*ill-health	x	x	(0.58)	x	x	x
Treatment*ethnicity	x	x	x	0.20	x	x
Treatment*ethnicity	x	x	x	(0.60)	x	x

Variables	Female	Age	Ill-health	Ethnicity	Employment status	Deprivation
Treatment*employed	x	x	x	x	-0.13	x
Treatment*employed	x	x	x	x	(0.43)	x
Treatment*deprivation	x	x	x	x	x	x
Treatment*deprivation	x	x	x	x	x	x
N	658	658	658	658	658	658
Constant	0.05	0.07	0.20	0.03	-0.01	-0.03
Constant	(0.58)	(0.57)	(0.64)	(0.59)	(0.60)	(0.58)
Provider variance	2.38	2.37	2.41	2.39	2.38	2.31
Provider variance	(1.50)	(1.50)	(1.51)	(1.52)	(1.49)	(1.47)

Note: Significance Levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors clustered at provider level, with region fixed effects. Standard errors presented in parathesis. Notation x indicates data were unavailable. Source: ILR Base: All learners, 719

Table A6 - 6: CACE analysis for primary and secondary outcomes

Variables	Primary outcome	Confidence in maths	Composite confidence
Treatment	0.01	-0.22	0.31
Treatment	(-0.25)	0.16)	0.26)
Age	0.00*	0.00	0.00
Age	0.00)	0.00)	0.00)
Female (reference: male)	0.01	-0.27*	0.43*
Female (reference: male)	0.06)	0.14)	0.18)
Ill-health (reference: no)	-0.02	-0.31**	-0.23
Ill-health (reference: no)	(0.07)	(0.11)	(0.18)
White (reference: others)	0.06	0.09	0.15
White (reference: others)	0.05)	0.09)	0.15)
Employed (reference: no)	0.04	-0.13	0.23*
Employed (reference: no)	(0.06)	(0.10)	(0.09)
Most deprived (reference: no)	-0.05	0.06	-0.39**
Most deprived (reference: no)	(0.05)	(0.16)	(0.14)
CA preference (= yes)	0.11	-0.16	-0.22
CA preference (= yes)	(0.10)	(0.20)	(0.29)
Baseline (confidence in maths)	x	0.26***	x
Baseline (confidence in maths)	x	(0.04)	x
Baseline (composite confidence)	x	x	0.28*
Baseline (composite confidence)	x	x	0.11
East of England	-0.13	0.02	-0.40***
East of England	(0.07)	(0.14)	0.11)
London	-0.09	0.13	0.03
London	(0.17)	(0.14)	0.20)
North East	-0.56***	x	x
North East	(0.07)	x	x
North West	0.02	0.19	0.19
North West	(0.11)	(0.14)	(0.46)
South East	0.11	0.28	-0.26
South East	(0.14)	(0.22)	(0.29)
South West	-0.06	x	x
South West	(0.14)	x	x
Yorkshire and the Humber	-0.07	-0.17	0.02
Yorkshire and the Humber	(0.19)	(0.10)	(0.19)
N	669	76	75
Constant	0.39***	3.56***	3.04***
Constant	(0.06)	(0.26)	(0.76)

Note: Significance Levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors clustered at provider level, with region fixed effects. Standard errors presented in parathesis. Notation x indicates data were

unavailable. Source: ILR Base: All learners for primary outcome and survey respondents for secondary outcome.

Table A6 - 7: Testing prediction of missing data (MAR)

Variables	Primary outcome	Confidence in maths	Composite confidence
Treatment	0.13	-0.75***	-0.86***
Treatment	(0.12)	(0.19)	(0.22)
Baseline (confidence in maths)	x	0.12	x
Baseline (confidence in maths)	x	(0.17)	x
Baseline (composite confidence)	x	x	0.10
Baseline (composite confidence)	x	x	(0.19)
Age	-0.01	-0.01	-0.01
Age	(0.01)	(0.01)	(0.01)
Female (reference: male)	-0.11	0.22	0.08
Female (reference: male)	(0.34)	(0.30)	(0.35)
Ill-health (reference: no)	0.24	0.15	0.24
Ill-health (reference: no)	(0.38)	(0.42)	(0.42)
White (reference: others)	0.50*	-0.12	-0.24
White (reference: others)	(0.23)	(0.60)	(0.51)
Employed (reference: no)	0.10	-0.42	-0.38
Employed (reference: no)	(0.23)	(0.43)	(0.45)
Most deprived (reference: no)	-0.07	0.09	0.05
Most deprived (reference: no)	(0.20)	(0.61)	(0.64)
CA preference (= yes)	-0.66***	-1.35***	-1.37***
CA preference (= yes)	(0.11)	(0.10)	(0.14)
East Midlands	0.23	-0.45	-0.44
East Midlands	(0.13)	(0.25)	(0.25)
East of England	0.81***	-0.58	-0.61
East of England	(0.14)	(0.56)	(0.55)
North East	x	x	x
North East	x	x	x
North West	0.69***	-0.94***	-0.95***
North West	(0.16)	(0.20)	(0.23)
South East	-2.13***	-0.75	-0.89
South East	(0.14)	(0.66)	(0.75)
South West	-0.57*	x	x
South West	0.28)	x	x
Yorkshire and the Humber	-0.20	-1.46***	-1.71***
Yorkshire and the Humber	(0.21)	(0.13)	(0.16)
N	575	163	161
Constant	-0.86	1.13	1.49
Constant	(0.50)	(1.01)	(1.02)

Note: Significance Levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors presented in parathesis. Notation x indicates data were unavailable. Source: ILR, baseline and endline survey.

Table A6 - 8: MICE results

Variables	Primary outcome	Confidence in maths	Composite confidence
Treatment	0.03	-0.02	0.07
Treatment	(0.13)	(0.12)	(0.15)
Baseline (confidence in maths)	x	0.08	x
Baseline (confidence in maths)	x	(0.06)	x
Baseline (composite confidence)	x	x	0.10
Baseline (composite confidence)	x	x	(0.08)
Age	0.00*	0.00	0.00
Age	(0.00)	(0.01)	(0.00)
Female (reference: male)	0.00	-0.06	0.14
Female (reference: male)	(0.06)	(0.16)	(0.14)
Ill-health (reference: no)	-0.04	-0.22*	-0.20*
Ill-health (reference: no)	(0.07)	(0.10)	(0.09)
White (reference: others)	0.03	-0.21	-0.09
White (reference: others)	(0.03)	(0.15)	(0.11)
Employed (reference: no)	0.00	0.01	0.02
Employed (reference: no)	(0.04)	(0.13)	(0.10)
Most deprived (reference: no)	0.00	0.03	-0.18
Most deprived (reference: no)	(0.03)	(0.12)	(0.10)
CA preference (= yes)	0.07	-0.07	0.13
CA preference (= yes)	0.11)	(0.12)	(0.13)
East of England	-0.29*	-0.10	-0.17
East of England	(0.14)	(0.12)	(0.11)
London	x	0.04	-0.19
London	x	(0.14)	(0.17)
North East	-0.70	0.48	0.47
North East	(0.11)	(0.33)	(0.30)
North West	-0.17	0.23	0.06
North West	(0.15)	(0.15)	(0.16)
South East	-0.10***	0.32	-0.08
South East	(0.21)	(0.17)	(0.19)
South West	-0.13	x	x
South West	(0.14)	x	x
Yorkshire and the Humber	-0.30	-0.32*	-0.12
Yorkshire and the Humber	(0.21)	(0.15)	(0.11)
N	719	260	260
Constant	0.50***	4.31***	4.05***

Variables	Primary outcome	Confidence in maths	Composite confidence
Constant	(0.09)	(0.32)	(0.36)
Provider variance (constant)	0.23	0.00	0.00
Provider variance (constant)	0.04)	.	(0.06)
Provider (residual)	0.46	0.75	0.67
Provider (residual)	(0.02)	(0.04)	(0.05)

Note: Significance Levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors presented in parathesis. Notation x indicates data were unavailable. Source: ILR, baseline and endline survey.

Table A6 - 9: Robustness checks

Variables	Model 1	Model 2	Model 3
Treatment	0.14	1.49	2.05
Treatment	(1.16)	(1.00)	(1.24)
Age	0.02**	0.02*	0.03
Age	(0.01)	(0.01)	(0.02)
Female (reference: male)	0.06	0.02	-0.07
Female (reference: male)	(0.33)	(0.27)	(0.28)
Ill-health (reference: no)	0.07	0.17	-0.03
Ill-health (reference: no)	(0.33)	(0.47)	(0.30)
White (reference: others)	0.19	0.23	1.09***
White (reference: others)	(0.17)	(0.22)	(0.32)
Employed (reference: no)	-0.02	0.11	0.02
Employed (reference: no)	(0.19)	(0.21)	(0.30)
Most deprived (reference: no)	0.01	0.01	0.17
Most deprived (reference: no)	(0.12)	(0.13)	(0.19)
CA preference (= yes)	0.93	0.91	0.17
CA preference (= yes)	(0.93)	(0.81)	(0.77)
East of England	-1.76	-0.91	-0.07
East of England	(1.21)	(0.98)	(0.54)
London	-0.89	-0.34	-0.34
London	(1.20)	(0.95)	(1.21)
North East	x	x	x
North East	x	x	x
North West	-0.19	0.41	0.43
North West	(0.75)	(0.66)	(0.66)
South East	-0.43	-0.59	-1.35
South East	(1.67)	(0.99)	(1.25)
South West	-0.79	-1.81*	-2.62*
South West	(0.97)	(0.80)	(1.19)
Yorkshire and the Humber	-1.82	0.31	-0.42
Yorkshire and the Humber	(1.40)	(1.06)	(1.37)
N	616	538	420

Variables	Model 1	Model 2	Model 3
Constant	-0.16	-0.67	-0.18
Constant	(0.65)	(0.47)	(0.56)
Provider variance (constant)	3.40	1.27	0.69
Provider variance (constant)	(2.18)	(1.20)	(0.57)

Note: Significance Levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors presented in parathesis.
Notation x indicates data were unavailable. Source: ILR.

Table A6 - 10: CACE analysis for FSQ course completion

Variables	Model 4 (Stage 1)	Model 2 (Stage 2)
Treatment	0.83	0.73
Treatment	(0.46)	(1.38)
Completion status FSQ	x	-5.25
Completion status FSQ	x	(6.98)
uhat	x	9.82
uhat	x	(7.21)
Age	0.01	0.04***
Age	(0.01)	(0.01)
Female (reference: male)	0.03	-0.03
Female (reference: male)	(0.22)	(0.40)
Ill-health (reference: no)	0.29	0.32
Ill-health (reference: no)	(0.50)	(0.38)
White (reference: others)	-0.46	0.07
White (reference: others)	(0.33)	(0.67)
Employed (reference: no)	0.03	0.02
Employed (reference: no)	(0.27)	(0.26)
Most deprived (reference: no)	0.01	0.06
Most deprived (reference: no)	(0.15)	(0.15)
CA preference (= yes)	0.61	1.64
CA preference (= yes)	(0.45)	(1.16)
East of England	0.37	-2.41
East of England	(0.65)	(1.43)
London	0.81	-0.95
London	(0.58)	(1.67)
North East	1.14***	x
North East	(0.32)	x
North West	0.54	-0.37
North West	(0.59)	(1.14)
South East	0.88	-0.46
South East	(0.77)	(2.03)
South West	-0.30	-1.62
South West	(0.45)	(1.24)
Yorkshire and the Humber	1.73**	-1.36

Variables	Model 4 (Stage 1)	Model 2 (Stage 2)
Yorkshire and the Humber	(0.62)	(2.48)
N	623	616
Constant	0.53	3.23
Constant	(0.49)	(4.83)
Provider variance (constant)	0.40	4.28
Provider variance (constant)	(0.28)	(2.42)

Note: Significance Levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors presented in parathesis.
Notation x indicates data were unavailable. Source: ILR.

Appendix 7: Additional IPE Tables

Table A7 - 1: Learner interview sample details

Variables	Variables	Count
Provider region (N=32)	East Midlands	3
Provider region (N=32)	East of England	2
Provider region (N=32)	London	13
Provider region (N=32)	North East	2
Provider region (N=32)	North West	2
Provider region (N=32)	South East	7
Provider region (N=32)	South West	1
Provider region (N=32)	Yorkshire & Humber	2
Gender (N=26)	Female	20
Gender (N=26)	Male	6
Age group (N=32)	Under 25	6
Age group (N=32)	25-49	20
Age group (N=32)	50 and over	6
Attendance (N=27)	Lower than 80%	10
Attendance (N=27)	80% or higher	17

Source: IPSOS learner data

Note that while 32 learners were interviewed, data for gender was available for 26 learners, and data for attendance was available for 27 learners. Note that for attendance, compliance is defined as attending at least 80% of sessions (i.e. 10 lessons or more)

© 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: RR1628/6
ISBN: 978-1-83870-796-5

For any enquiries regarding this publication, contact www.gov.uk/contact-dfe.

This document is available for download at www.gov.uk/government/publications.