PROJECT TITLE	Evaluation of the Online Professional Development for Delivering the Mathematical Reasoning Programme
DEVELOPER (INSTITUTION)	University of Oxford
EVALUATOR (INSTITUTION)	University College London
PRINCIPAL INVESTIGATOR(S)	Dr. Tamjid Mujtaba (PI) and Dr. Richard Sheldrake (Co I)
EVALUATION PLAN AUTHOR(S)	Dr. Tamjid Mujtaba and Dr. Richard Sheldrake
PUPIL AGE RANGE AND KEY STAGE	Year 2 (aged 6 to 7 years old; Key Stage 1)
NUMBER OF SCHOOLS/ SETTINGS	30+ schools (expected)
NUMBER OF TEACHERS[1]	At least one teacher per school (30+ expected)

Evaluation plan version history

VERSION	DATE	REASON FOR REVISION
1.2 [*latest*]		
1.1		
1.0 [*original*]	28/02/2023	*[leave blank for the original version]*

[1] Adapt as needed to relevant sample for the pilot (e.g., number of teachers)

Intervention

Background and context

A range of work has informed the approaches and foci of the contemporary Mathematical Reasoning programme. Specifically, the developers of the Mathematical Reasoning programme have previously produced and delivered training for pupils around understanding numbers (including the inverse relation between addition and subtraction) which was associated with improvements in the pupils' mathematics test scores (Nunes, et al., 2007). The developers have also undertaken research that has associated indicators of mathematical reasoning and arithmetic with pupils' subsequent attainment in mathematics (Nunes, Bryant, Barros, & Sylva, 2012).

The Mathematical Reasoning programme has been previously evaluated through an efficacy trial under the name of Mathematics and Reasoning (Worth, Sizmur, Ager, & Styles, 2015). The efficacy trial involved 17 allocated intervention schools receiving Mathematics and Reasoning as a numeracy programme and 19 allocated intervention schools receiving Literacy and Morphemes as a literary programme, compared to 19 allocated control schools; the Mathematics and Reasoning programme was associated with a positive effect on mathematics test scores (an effect size of 0.20) compared to control schools, while the Literacy and Morphemes programme was associated with a negative effect on English test scores (an effect size of -0.05) compared to the same control schools (Worth, Sizmur, Ager, & Styles, 2015).

The Mathematical Reasoning programme has also been previously evaluated through an effectiveness trial (Stokes, et al., 2018). The effectiveness trial involved 84 allocated intervention schools receiving Mathematical Reasoning compared to 85 allocated control schools; the Mathematical Reasoning programme was associated with a positive effect on mathematics test scores (an effect size of 0.08) compared to control schools, although the effect was not statistically significant (Stokes, et al., 2018).

The Mathematical Reasoning programme has now been adapted into an online training format. The online professional development essentially aims to support and facilitate teachers to deliver the Mathematical Reasoning programme, together with other support provided through the programme. This evaluation focuses on the online professional development and considers various areas including whether/how teachers felt that the training was effective and acceptable, and whether any refinements may be beneficial.

Mathematical Reasoning training and delivery approaches

The provision of training and support within the Mathematical Reasoning programme has developed over time.

The Mathematics and Reasoning programme, as evaluated within the efficacy trial, involved one day of initial training for participating teachers (delivered in-person by the developers), one subsequent visit to schools (undertaken by the developers in order to help support the implementation of the programme), and provided teaching materials to be applied by the participating teachers.

The Mathematical Reasoning programme, as evaluated within the effectiveness trial, involved the developers training 'Work Group Leads' (through one day of training delivered in-person), which were teachers who then delivered the programme with their own pupils; the 'Work Group Leads' then received further training (through two days of training delivered in-person) in order to support them to train other teachers. The 'Work Group Leads' then provided training to the participating teachers (through one day of training delivered in-person) and undertook one subsequent visit to participating schools in order to help support the implementation of the programme (the visit included, where possible, a lesson observation and provided teachers

with the opportunity to discuss any issues they were facing in implementing the programme). Teaching materials were provided, which were applied by the participating teachers, and further online support was also available to the participating teachers.

The programme developers have provided more details about the training and delivery approaches (Appendix A).

The Mathematical Reasoning training has now been adapted into an online format, where teachers (and teaching assistants as/if applicable) are expected to complete a series of modules covering the principles and approaches of the programme. 'Work Group Leads' are now referred to as 'Teacher Leaders', and the role has adapted such that the Teacher Leaders essentially provide online support to teachers and teaching assistants in order to help the implementation of the programme.

The Mathematical Reasoning programme continues to aim to promote pupils' understanding of numbers or number sense (encompassing additive composition of number and place value, and the inverse relation between addition and subtraction) and understanding of quantitative reasoning or problem solving (different relations that can be established between quantities, encompassing part whole or additive relations, and one to many correspondence or multiplicative relations).

Template for Intervention Description and Replication (TIDieR): Mathematical Reasoning via online training

1. Brief name

Mathematical Reasoning

2. Why (rationale/theory)

Pupils' understanding of numbers and quantitative reasoning may be beneficial within mathematics education (Nunes, Bryant, Barros, & Sylva, 2012; Nunes, et al., 2007). Specifically, previous research has shown that quantitative reasoning and arithmetic have associated with students' subsequent mathematical attainment, with quantitative reasoning having the stronger association (Nunes, Bryant, Barros, & Sylva, 2012); an intervention with children who performed poorly in a quantitative reasoning assessment was effective in improving the children's quantitative reasoning and had a significant impact on their Key Stage 1 mathematics assessment, taken a year after the intervention (Nunes, et al., 2007).

The designers of the Mathematical Reasoning programme have carried out a number of studies to investigate how different aspects of children's quantitative reasoning develops (Nunes, Bryant, Evans, & Bell, 2010; Nunes, Bryant, Evans, Bell, & Barros, 2012; Nunes, Bryant, Hallett, Bell, & Evans, 2009; Nunes, et al., 2009) and have created activities that were found to be effective in promoting pupils' progress on four concepts that are critical for mathematical development: the additive composition of number; the inverse relation between addition and subtraction; the understanding part-whole; and the understanding of ratio relations between quantities. These activities were put together into the Mathematical Reasoning programme (initially called Mathematics and Reasoning in the efficacy trial), which was found to have a significant positive impact (magnitude = 0.20, p = 0.03) on children's mathematics performance in a standardised test through an efficacy trial (Worth, Sizmur, Ager, & Styles, 2015) and a non-significant smaller positive impact (magnitude = 0.08, p = 0.16) through an effectiveness trial (Stokes, et al., 2018).

In the efficacy trial of the Mathematical Reasoning programme, teachers were directly trained by the programme designers; the subsequent effectiveness trial involved the programme designers training 'Work Group Leads' who then provided training to the participating teachers.

The current pilot evaluation aims to evaluate the effectiveness of training teachers through an online course created by the programme designers, which can inform the feasibility of further

and wider delivery. If the online course is effective in supporting the implementation in the classrooms, teachers in the future could have access to training without the restrictions of staff, time, and location that come from direct face-to-face training by the programme designers.

3. Who (recipients)

The recipients of the Mathematical Reasoning programme are Year 2 pupils in the classes of teachers (and teaching assistants if applicable) who have had the training; the programme engages with one class per school. Teaching assistants are expected to undertake the training with the teachers in order to discuss the activities and then deliver them to Year 2 pupils.

The programme involves:

- Year 2 pupils, who receive the programme;
- Year 2 teachers and teaching assistants, who undertake online training and who deliver the programme (using the following materials and procedure);
- Teacher Leaders, who are trained to provide personalised support to the Year 2 teachers and teaching assistants (as per the following procedures);
- The Mathematical Reasoning team, who train the Teacher Leaders and provide further organisational support (as per the following procedures).

4. What (materials)

Teachers (and teaching assistants if applicable) are provided online training, teaching materials and supporting resources, and online games for pupils. 'Teacher Leaders' (also referred to as a workgroup/course leaders) are trained to provide personalised support via a website forum and also provide 3 online webinars/tutorials for teachers and teaching assistants.

Online training for teachers involves:

- videos of training led by the developers, organised into 'modules' for teachers to work through;
- videos of work carried out with children, to model the activities that teachers can use in the classroom;
- reflection activities and tasks to show their understanding;
- downloadable research brief documents for teachers, which offer further/supplementary information and opportunities for reflection, and downloadable presentation slides as used within the video training led by the developers.

Teaching materials and supporting resources involve:

- A teaching handbook with a unit plan and instructions for each activity;
- Teaching presentations for 12 units;
- A children's workbook (written activities for children to complete during the teacher-led parts of each unit);
- Online games for children;
- A children's record sheet (for children to tick the online games played per unit);
- Supplementary physical games (if children need extra support before they access the independent computer games);

- Games that use manipulatives including cards;
- Instructions to use the website, and a 'frequently asked questions' document.

An online website forum for teachers provides an avenue for personalised support via written questions and discussions with Teacher Leaders.

Teacher Leaders provide personalised support through 3 online webinars/tutorials with teachers (and teaching assistants if applicable).

5. What (procedures)

Teacher Leaders:

- Undertake two days of training from the Mathematical Reasoning team;
- Book dates for online tutorials with teachers;
- Monitor the discussions that are part of the course and approve the teachers' access to the website after the teacher has completed four modules;
- Monitor an 'online forum' (i.e. the Chat in Canvas) where teachers can post questions about the programme and its delivery and answer questions once a week;
- Hold three one-hour webinars with teachers (and teaching assistants if applicable) in the groups of schools under their responsibility.

The Mathematical Reasoning team:

- Train the Teacher Leaders;
- Register schools on the online course;
- Send printed materials to schools;
- Monitors the discussions in the online course to identify aspects that require further attention or modification;
- Provide guidance and support for the Teachers Leaders.

Teachers (and teaching assistants if applicable):

- Undertake online training modules 1-4 during one day of professional development; completion of the first four modules gives the access to the website with the programme materials, including whole class teaching materials, and online games;
- Undertake online training module 5 (covering practical guidance for implementing the programme) before delivery with students;
- Are encouraged to undertake online training modules 6-9 flexibly, while they are delivering the programme to their classes.

Teachers deliver 12 teaching units (one hour each) involving:

- Teacher-led activities (whole class activities);
    - Demonstrations with objects ('manipulatives for demonstrations and for enacting story problems');
    - Presentations where students solve problems

- Group activities with students separated into Level 1 and Level 2 groups, depending on how they react of the content of the lesson (Level 1 activities are for children who need some more time and support with learning new concepts; Level 2 activities are for children who are quicker to grasp new ideas and can be helped to extend their thinking):
    - One group focuses on consolidation of what has been taught in the lesson through online games;
    - The other group has more intensive, differentiated teaching with the teacher which might be an extension to greater depth, some extra support with a concept, or some pre-teaching to support the next lesson.

6. Who (implementers)

Teachers (and teaching assistants if applicable) who receive training. 'Teacher Leaders' are recruited, and also receive training for their role, by the Mathematical Reasoning team.

7. How (mode of delivery)

During normal mathematics lessons; lessons from the programme are intended to replace the (National Curriculum based) maths lessons that the teachers would otherwise be giving (so no additional maths teaching is required).

8. Where (setting)

Regular classrooms in participating schools.

9. When and how much (dosage)

Teachers deliver 12 teaching units (one hour each) with their pupils, with one unit undertaken per week (where a 'unit' can also be referred to as a 'lesson' and which encompass teacher-led activities, groups activities, and online games).

Online training modules 1-5 are intended to be compulsory (and undertaken before delivery with students); modules 6-9 are considered optional.

Teachers are expected to attend 3 online webinars/tutorials with the Teacher Leader responsible for their group.

10. Tailoring

Teachers are expected to apply all units in order, and apply all activities as planned and structured per unit.

The training and handbook recommend that the structure of the unit plans is followed as written; if a unit is not completed in one lesson, the handbook recommends resuming at the place stopped in the previous lesson so that all material is covered.

The training explains to teachers that fidelity is the extent to which the programme is delivered as intended, and adaptations could result in not achieving the impact that may be possible if a programme were to be implemented faithfully with high fidelity.

Teachers are not explicitly told that adaptations are impossible but are recommended to start with an assumption that the programme should be delivered as intended. The training conveys that if anything appears problematic, then it may be because the programme team have not made it clear enough (in guidelines or lectures); that before making an adaptation, teachers are recommended to ask questions in the online forum so that Teachers Leaders can support and clarify to avoid miscommunications. (Teacher Leaders are expected to handle questions once a week.)

Theory of change / logic model

The logic model for the training as theorised and proposed by the delivery team (Appendices B and C) essentially involves:

- Resources and inputs including the online training modules and programme handbook;
- Processes of teachers undertaking the training (and then applying the programme activities and materials in particular ways with their pupils, all of which entails compliance with the programme as intended);
- Expected outcomes of teachers having confidence and knowledge around the underlying mathematical aspects and principles underlying the delivery of the programme.

Some of these elements also occur within a wider programme logic model, essentially where changes to usual practices (the application of programme activities and materials in particular ways, all of which entails compliance with the programme as intended) are theorised to entail benefits to pupils (understanding of numbers or number sense and understanding of quantitative reasoning or problem solving) that may be observable through mathematics test scores (and/or potentially through wider examination outcomes within education).

On a training level, teacher and school aspects (such as the teachers' existing knowledge, whether the training is accessible to teachers, etc.) may act as moderating factors that may influence the magnitude of any effect on teacher-level outcomes.

On a programme level, undertaking the training (and other aspects of compliance with the programme as intended) and/or the training outcomes themselves (teachers having confidence and knowledge around the underlying mathematical aspects and principles) may act as moderating factors that may influence the magnitude of impact on any pupil-level outcomes. Any teacher-level aspects (such as the teachers' existing knowledge) may also act as moderating factors that may influence the magnitude of impact on any pupil-level outcomes. Wider school-level aspects (such as teachers having support, flexibility, and/or time) may also be relevant as moderating factors, which may facilitate or contain engagement with the training and/or any application of the programme as intended.

On a programme level, training outcomes such as teachers having confidence and knowledge around the underlying mathematical aspects and principles (and/or other aspects such as features of the programme being perceived as interesting and enjoyable by pupils) may potentially act as mediating factors that explain any pupil-level outcomes. This would essentially involve pupil-level outcomes occurring because teachers have gained confidence and knowledge around the underlying mathematical aspects and principles (which facilitates teachers to convey understanding to pupils).

On a programme level, the extent of adaptation (fidelity of implementation) may also act as a moderating factor that may influence the magnitude of impact on any pupil-level outcomes.

The logic modelling and theorising involves some wider assumptions (and potential uncertainty). The programme is assumed to replace the mathematics lessons that the teachers would otherwise be giving (so that schools applying the programme and any other schools are assumed to spend the same amount of time on mathematics teaching/learning); nevertheless, the time spent on mathematics teaching/learning may act as a mediating factor that associates with any pupil-level outcomes. Additionally, the programme involves an

extensive array of activities where 12 teaching units are intended to be applied in a particular order, where each unit involves teacher-led activities, group activities (for different groups of pupils), and also online games.

Research questions

This research is a pilot evaluation of the online training within the Mathematical Reasoning programme, focused on acceptability, implementation, and effectiveness in preparing teachers.

Pilot evaluations often consider whether programmes are feasible to be implemented as intended, which may involve considering facilitators or barriers around particular programme designs and/or features, considering whether/how aspects within a theory of change occur and operate, and other areas (Education Endowment Foundation, 2022). More extensive efficacy and effectiveness trials then apply impact evaluations in order to quantify the magnitude of change or effect that can be inferred to be associated with a programme (Education Endowment Foundation, 2018), and apply implementation and process evaluations in order to understand aspects around the delivery of a programme such as whether/how a programme was delivered as intended (Education Endowment Foundation, 2019; Humphrey, Lendrum, Ashworth, Frearson, & Buck, 2016).

In more detail, pilot evaluations often consider three broad areas

- Evidence of promise broadly considers whether evidence supports a programme's theory of change, and may include considering actual or perceived impact, whether/how links between elements in a theory of change indeed occur, and/or the extent to which a programme differs from usual practice (Education Endowment Foundation, 2022).
- Feasibility of implementation broadly considers whether a programme can be delivered as intended, and can include aspects often seen within implementation and process evaluations such as the perceived acceptability of a programme, whether/how adaptations are made to a programme in practice, and what factors support or hinder the application of a programme (Education Endowment Foundation, 2022).
- Readiness for trial (or scalability) considers whether a programme can be applied on larger scales (such as across around 50 or more schools) and associated aspects such as the feasibility of reproducing a programme (including whether programme elements are identified and documented and whether programme materials are codified), and can consider costs and/or resource implications and aspects around eligibility criteria to receive a programme (Education Endowment Foundation, 2022).

The pilot evaluation of the online training of Mathematical Reasoning will consider the following research questions.

- Evidence of promise
    - RQ1. How effective is the online training and support at preparing teachers to deliver Mathematical Reasoning to their pupils?
        - RQ1.1 What are teachers' general perceptions of the online training and support?

        - RQ1.2 How effective is the online training at supporting teachers' confidence to deliver the programme (including their understanding of the underlying mathematical areas and programme principles)?
        - RQ1.3. To what extent did the teachers implement the programme as intended, and/or to what extent were they able to adapt the programme to their context (and how did this follow from the training)?

- Feasibility of Implementation
    - RQ2. Was the online training acceptable to schools and teachers? (Cost, burden on time, etc).
    - RQ3. How much of the training did teachers undertake (and why)? What are the enablers and barriers to engagement?
    - RQ4. Did the Teacher Leaders feel equipped to support the teachers, and to what extent did the teacher leaders have to adapt their support? What are the enablers and barriers?
- Readiness for trial (or scalability)
    - RQ5. Would the online training benefit from any changes or refinements before it could be applied on a larger scale?

Table 1: Success indicators for research questions and pilot criteria

Pilot criteria	Success Indicators	How to assess this?
Evidence of promise	RQ1: consensus of effectiveness	Questionnaire and interview responses; observations
Feasibility of implementation	RQ2: consensus of acceptability RQ3: majority of schools completing core training modules RQ4: absence of fundamental issues	Questionnaire and interview responses Reported and/or recorded completion of core training modules
Readiness for trial	RQ5: consensus of acceptability / absence of fundamental issues or barriers; suggested mitigation of issues/barriers and/or other suggested refinements/changes	Questionnaire and interview responses

Methods

Recruitment

The evaluation expects to recruit 32 schools. The University of Oxford will recruit 4 'Teacher Leaders'. Each Teacher Leader will then be expected to recruit 8 schools.

Schools will be provided with an information sheet explaining the programme and the evaluation activities, eligibility criteria, and related matters including data protection. Schools will then complete a Memorandum of Understanding in order to be included within the evaluation. Eligibility criteria are:

- The schools are located in West Yorkshire, South Yorkshire, or Merseyside;
- The schools have not used the Oxford Improving Working Memory Plus Arithmetic (IWM+A) intervention during the 2021/2022 academic year.

The sample size was determined by the potential for coverage by the delivery team and resources for data collection. The expected sample size is somewhat comparable to earlier implementation and process evaluations. The earlier efficacy trial included an implementation and process evaluation that considered questionnaire responses from 22 teachers who delivered Mathematics and Reasoning (Worth, Sizmur, Ager, & Styles, 2015) while the earlier effectiveness trial included an implementation and process evaluation that considered questionnaire responses from 50 teachers who delivered Mathematical Reasoning (Stokes, et al., 2018).

Measures

The evaluation will devise and apply measures within data collection (such as questionnaire items and interview prompts) in order to address the research questions within the context of the programme and training. The evaluation will focus on areas such as teachers' reported confidence and teachers' perceptions around the usefulness of the training elements, rather than formalised testing of skills or knowledge.

The first interview schedule will focus on experiences of the training and accessibility of materials. The second interview schedule will focus on experiences of the implementation of the programme. The survey will look at overarching views about the content of the training modules. Analysis of one element of data will help to inform the development of other parts of data collection, such as any available observations of the Teacher Leader training being used to inform the development of the questionnaire for teachers and teaching assistants, and/or responses from the questionnaire being used to inform the development of interview schedules.

The purpose of the initial survey is to gain an initial understanding of how the compulsory training modules are received by teachers, gathering experiences including teachers' perceptions of any barriers and views around what aspects were considered to be more helpful or less helpful. The survey will also collect data around how confident teachers feel after undertaking the training, and therefore provide a perspective onto the effectiveness of the training. The survey will also gather information about whether any changes to the training (as suggested by teachers) may be helpful.

Analysis of the teacher survey data and online forum data triangulated with the observation data of training and Mathematical Reasoning lessons will lead to the development of a teacher interview schedule where we will interview all 30 Year 2 teachers and their teaching assistants. Questions will include their views on training, implementation, pupil engagement and impact. The first set of interviews will be conducted soon after the training has been completed and will focus on just the training and accessibility of materials. The second round of interviews will focus on experiences of teachers in implementing the programme.

Data collection

Table 2: Summary of research methods

IPE dimension	RQ addressed	Research methods	Data collection methods	Sample size and sampling criteria	Data analysis methods
Evidence of promise	RQ 1.1-1.3	**Interviews**	Structured interviews	All sample (around 30 schools)	Frequency counts; deductive coding;
		Observations	Structured observations	10 schools	Frequency counts; deductive coding
		Surveys (post CPD and some lesson delivery)	Online questionnaires	30 teachers 30 TAs	Descriptive statistics; correlations; bi-variate analysis e.g., t-tests
Feasibility of implementation	RQ 2-4	**Interviews**	Structured interviews	All sample (around 30 schools)	Frequency counts; deductive coding;
		Surveys (post CPD and some lesson delivery)	Online questionnaires Observations	Teacher Leaders 30 teachers 30 TAs	Descriptive statistics; correlations; bi-variate analysis e.g., t-tests
Readiness for trial	RQ 5-6	**Surveys** (post CPD and some lesson delivery)	Online questionnaires	All sample (around 30 schools)	Frequency counts; deductive coding;
		Interviews	structured interviews	30 teachers 30 TAs	Descriptive statistics; correlations; bi-variate analysis e.g., t-tests

Overall design and structure

Mixed-methods approaches will be undertaken, covering interviews, surveys, and observations. The data collection will also, if/where possible, encompass wider programme elements such as online engagement and content from synchronous forums (essentially across the training modules and the online forum where teachers and teaching assistants engage with Teacher Leaders, such as modules being completed, time taken for completion, what questions are asked by the teachers, etc.).

The questionnaire will essentially cover teachers' perceptions of the initial (compulsory) training modules, relatively shortly after these have been completed in order to gather views

and experiences. However, at that time, teachers may not have started applying the programme (or may not have extensively applied the programme) with their pupils. This design nevertheless aims to ensure that teachers' views about the compulsory training content can be gathered, even if teachers might not participate within subsequent interviews. The observations will focus on the confidence, accuracy and engagement with pupils.

Once the teachers have started applying the programme, interviews will then be undertaken to gather teachers' detailed perceptions around the training in the context of facilitating them to apply the programme (encompassing wider areas such as teachers' understanding of the aims of the programme, whether/how the training has supported their delivery of the programme, etc.). This design applies interviews in order to effectively gather more detailed views from teachers, which may be difficult to gather through a questionnaire.

The initial training of the Teacher Leaders, and the webinars delivered by the Teacher Leaders, will also be observed if/as necessary. Final follow-up interviews will be conducted with the Teacher Leaders if/as necessary.

Participant numbers

Given the expected sample size and the evaluation focus, the questionnaire and interview coverage will aim to encompass all schools (expected to be around 30 schools). Given this is a small pilot study it would be useful to interview all teachers. This will enable us to capture responses from teachers from a range of professional and contextual backgrounds.

The sample cannot be assumed to be homogeneous, given the expected coverage of schools across multiple geographical regions and where teachers may have different characteristics (such as age, gender, and background circumstances) and where any characteristics would be unknown prior to data collection. It is possible that some characteristics/backgrounds may be relevant to perceptions of training, for example potentially where those with specialisms in mathematics may find the training easier to engage with (or perhaps harder to engage with).

Questionnaires and interviews can be undertaken remotely and only incur time costs, without incurring travel and other expense costs. Additionally, teachers may not necessarily choose to participate within all data collection activities, which can only be mitigated (to some extent) by approaching everyone in the sample. Essentially, there is no clear positive benefit that might follow from limiting questionnaire and interview data collection. Limiting data collection could introduce risk through the research inherently involving less data and less potential for insights, while also complicating consideration of whether/how a sample or sub-sample may generalise to wider populations and contexts.

Observations are more resource intensive than questionnaires and interviews. Observations will be undertaken through school visits (if this is possible/feasible for schools) or through remote video (with a video in a fixed location placed in front of teachers' desks facing away from pupils; observation notes will be taken but the video feed will not be recorded). Given the expected sample size, focus, and resourcing, teaching will be observed within a smaller number of schools (expected to involve 10 schools) to balance the potential to gather insights with the burden on teachers/schools and the available resources.

Data collection coverage and content

The questionnaires are expected to encompass areas such as the following.

- Views around the initial (compulsory) training modules and documentation: what particular modules and detailed elements are or are not engaged with (and why), perceived accessibility and support (and suggestions for any improvements), and teachers' initial confidence and other views around the covered topics and content.

The interviews are expected to encompass areas such as the following.

- More detailed views around the (optional) training modules and documentation (and/or any updated views around the initial modules now that the teachers will have had more experience in applying the programme): what particular modules and detailed elements are or are not engaged with (and why), perceived accessibility and support (and suggestions for any improvements), which programme elements are or are not applied (or applied with adaptation) within teaching (and why), and teachers' confidence and other views around the covered topics and content.

- Views around the wider training programme: the use, perceived acceptability, and other views around particular programme elements (e.g. aspects of the programme structure including the professional development day; programme elements such provision of feedback and the online forum; the instrument for teachers to assess children's learning, the reflection activities and tasks for teachers, usefulness of Teacher Leaders in terms of support, guidance and enabling accessibility etc.).

- Similarities/differences between the programme content and usual content and teaching.

- Teacher views about any perceived benefits/impacts in pupils.

- Perceived overall acceptability of the programme, and within wider frameworks and standards such as Ofsted.

- Perceptions and/or suggestions around wider provision of the training, such as which parts might be more or less feasible within potential up-scaling.

The observations are expected to encompass areas such as the following.

- Whether/how programme aims are clear/understood;

- Whether/how programme practices are as conveyed (and/or adapted);

- Confidence and other aspects of delivery (such as support and responses to pupils, whether any challenges are observable).

Data analysis

Quantitative data is expected to be analysed through summaries of frequencies and averages, given that the numbers of participants and scope of data are likely to be insufficient for more complex approaches and given that more complex approaches are not necessary to answer the research questions.

Qualitative data is expected to be analysed through identifying common points and themes conveyed through free-text responses and interview narratives, through iterative development

(e.g. multiple passes through and across data to identity potential themes and their refinement). Coding and themes (via a thematic approach) will also be guided by analysis of other sources of data (e.g. the initial survey, classroom observation etc.). Classroom observations will help address whether the teachers understood the purpose of the training and were able to develop the lessons as per guidance/training. The questionnaire and interviews are expected to cover teachers conveying aspects and accounts of what happens on the programme, and how the training facilitated their confidence and delivery, and the observations provide a further complementary perspective onto what happens on the programme as seen by others.

Triangulation will ensure that detailed insights can be gained through qualitative interviews supported by quantitative surveying and taking a theory generation theory verification approach, including using one strand of data collection to inform instrument development in another. Integration across methods will be facilitated through the questionnaires and interviews including some common or equivalent content (such as teachers' confidence in delivering aspects of the programme), where the qualitative methods will also allow more detailed responses from teachers.

The findings will also be considered within the context of the logic model, for example considering whether/how the online training facilitates teachers to feel confident in applying the programme (a theorised outcome within the logic model) and considering any relevant facilitating factors and/or barriers (which might act as moderating factors within the logic model). However, evaluation does not aim to test any proposed mechanisms or features within the logic model, which would require a more extensive trial, sample, and/or different approaches (such as also including data collection with pupils).

Ethics and registration

The research will be reviewed and approved by the IOE ethics committee before data collection commences.

Data protection

Processing any personal data will be undertaken in accordance with General Data Protection Regulation (GDPR) principles under the 'public task' legal basis.

UCL will be the controller of the data collected for the evaluation. All data collected through the evaluation will be kept secure and confidential.

All responses will be anonymised, and participants and schools will not be identifiable within any reporting of results. Results may include short quotations from interviews, where no-one can be identified.

After findings are disseminated, data will be anonymised and stored securely for 10 years.

After the evaluation is complete, data will be retained for research purposes such as presentations at professional or academic conferences or publications in professional or academic journals.

Personnel

The Mathematical Reasoning delivery team includes the following.

- Gabriel Stylianides (University of Oxford), PI, co-designed the online training and contributed to the design and implementation of training of TLs.
- Terezinha Nunes (University of Oxford) co-designed the online training and was part of the initial team that developed the Mathematical Reasoning programme; contributed to the design and implementation of training of TLs.
- Louise Matthews (University of Oxford) contributed to the development of the online professional development training and to the design and training of TLs; coordinates the training of TLs and programme delivery as well as school recruitment.
- Teacher Leaders contributed to school recruitment and support the participating teachers that deliver the programme.

The evaluation team includes the following:

- Tamjid Mujtaba (IOE, UCL's Faculty of Education and Society) is the Principal Investigator with responsibility for the overall evaluation and contributes research and evaluation expertise.
- Richard Sheldrake (IOE, UCL's Faculty of Education and Society) contributes research expertise and quantitative evaluation aspects.
- Eirini Geraniou (IOE, UCL's Faculty of Education and Society) contributes mathematical and research expertise.
- Andy Markwick (IOE, UCL's Faculty of Education and Society) contributes mathematical and research expertise.
- UCL research assistants and academics to help with the qualitative element.

Risks

Key risks involve recruitment and engagement that determine the overall sample size and the extent of data that may be collected within the evaluation, and the relatively short timeframe where the evaluation is essentially undertaken within one year (which may limit attempts to maximise recruitment and data collection, and may limit the overall extent and complexity of work that can be feasibly undertaken).

Recruitment determines the overall sample size. Covering around 30 schools limits the extent of data that can be collected; data collection may also be limited by varying engagement with different evaluation activities (such as teachers engaging with questionnaires but not interviews) and non-response within data collection (such as teachers not answering some questions within a questionnaire). Nevertheless, the earlier Mathematical Reasoning trials involved some implementation and process evaluation aspects considering responses from around 20-30 teachers, which helps affirm the plausibility of insights emerging from relatively small samples. The potential for insight will be maximised through including free-text questions within questionnaires and through in-depth interviewing, essentially intending to mitigate some risk by focusing on depth and detail within data collection.

Recruitment of schools is planned between December 2022 through January 2023. Schools are planned to deliver Mathematical Reasoning from February 2023 through May / mid-June 2023. External factors (such as the COVID-19 pandemic, teacher strikes, teachers attending to other matters within their schools, etc.) may impact the delivery of the programme within schools and may also limit data collection. Schools may drop out of the study, and recruiting replacement schools may not be possible within a limited timeframe. Some of this risk is intended to be mitigated through concurrent collection and analysis of data, although an evaluation cannot realistically mitigate against all external risks.

Timeline

Table 3: Timetable including specification of who completes each task

Dates	Activity	Staff responsible/ leading
December 2022	Recruitment of Teacher Leaders	University of Oxford
December 2022 / January 2023	Training of Teacher Leaders (14 December 2022 and 17 January 2023)	University of Oxford
December 2022 / January 2023	Recruitment of schools	Teacher Leaders
January-February 2023	Teachers undertake first 5 online training modules prior to delivering any Mathematical Reasoning activities	Teachers in participating schools
February-April 2023	Survey of teachers (perceptions/experiences of initial training)	UCL
February-June 2023	Teachers deliver the Mathematical Reasoning programme while undertaking further online training modules	Teachers in participating schools
February-June 2023	Three online tutorials with teachers	Teacher Leaders
April-May 2023	Interviews of teachers (mid-way through programme delivery; perceptions/experiences of training in the context of programme delivery)	UCL
February-June 2023	Observations of 10 lessons	UCL
19th June 2023	Presentation of interim findings to EEF	UCL
June-October 2023	Interviews of teachers (end of programme delivery; perceptions/experiences of training in the context of programme delivery)	UCL
August-September 2023	Draft report submitted to EEF	UCL
March - April 2024	Final report published and anonymisation of data	UCL

References

Education Endowment Foundation. (2018). *Statistical analysis guidance for EEF evaluations [March 2018].* London: Education Endowment Foundation.

Education Endowment Foundation. (2019). *Implementation and process evaluation guidance for EEF evaluations, August 2019.* London: Education Endowment Foundation.

Education Endowment Foundation. (2022). *Guidance for pilot evaluations.* London: Education Endowment Foundation.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., & Buck, R. (2016). *Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook.* London: Education Endowment Foundation.

Nunes, T., Bryant, P., Barros, R., & Sylva, K. (2012). The relative importance of two different mathematical abilities to mathematical achievement. *British Journal of Educational Psychology, 82*(1), 136-156. doi:10.1111/j.2044-8279.2011.02033.x

Nunes, T., Bryant, P., Burman, D., Bell, D., Evans, D., & Hallett, D. (2009). Deaf Children's Informal Knowledge of Multiplicative Reasoning. *Journal of Deaf Studies and Deaf Education, 14*(2), 260-277. doi:10.1093/deafed/enn040

Nunes, T., Bryant, P., Evans, D., & Bell, D. (2010). The scheme of correspondence and its role in children's mathematics. *British Journal of Educational Psychology, 2*(7), 83-99. doi:10.1348/97818543370009X12583699332537

Nunes, T., Bryant, P., Evans, D., Bell, D., & Barros, R. (2012). Teaching children how to include the inversion principle in their reasoning about quantitative relations. *Educational Studies in Mathematics, 79*, 371-388. doi:10.1007/s10649-011-9314-5

Nunes, T., Bryant, P., Evans, D., Bell, D., Gardner, S., Gardner, A., & Carraher, J. (2007). The contribution of logical reasoning to the learning of mathematics in primary school. *British Journal of Developmental Psychology, 25*(1), 147-166. doi:10.1348/026151006X153127

Nunes, T., Bryant, P., Hallett, D., Bell, D., & Evans, D. (2009). Teaching Children About the Inverse Relation Between Addition and Subtraction. *Mathematical Thinking and Learning, 11*(1-2), 61-78. doi:10.1080/10986060802583980

Stokes, L., Hudson-Sharp, N., Dorsett, R., Rolfe, H., Anders, J., George, A., . . . Munro-Lott, N. (2018). *Mathematical Reasoning Evaluation report and executive summary.* London: Education Endowment Foundation.

Worth, J., Sizmur, J., Ager, R., & Styles, B. (2015). *Improving Numeracy and Literacy Evaluation report and Executive summary.* London: Education Endowment Foundation.

Appendix A: Mathematical Reasoning training and delivery approaches

The programme developers have provided the following details.

In the efficacy trial, professional development offered to teachers to implement the Mathematical Reasoning programme involved one day of initial training delivered in-person by the developers, one subsequent visit to schools by the developers and the opportunity to ask questions by contacting the designers by phone or email. Teachers received a Teacher Handbook and were provided with teaching materials to be used in the classroom, which included PowerPoint presentations for delivering whole class activities and PowerPoint files to be printed as children's record sheets for the whole class activities. Feedback from teachers indicated that they had spent considerable amounts of time printing the children's record sheets for use in the classroom, which also had cost implications. The delivery team then created a Child Workbook, which includes all the pages that teachers previously printed for use in the classroom.

In the effectiveness trial, professional development was based on the train-the-trainer model. The programme developers trained 'Work Group Leads' recruited by the NCETM (National Centre for Excellence in Teaching Mathematics). Work Group Leads received initially the same training that had been offered to teachers in the efficacy trial to prepare them to deliver the programme to their own pupils. As in the efficacy trial, they received a Teacher Handbook to support their delivery of the programme and had access to all the materials previously made available to teachers. After completing delivery to one class, 'Work Group Leads' received two further days of training delivered in-person by a leader from the NCETM with support from the delivery team. During these two days, Work Group Leads presented examples of work from their own classes, discussed the challenges and solutions they had noted during delivery, and used the PowerPoint slides that the delivery team had created for training teachers as part of their preparation to train teachers. The programme designers had no further influence on the training of teachers. The 'Work Group Leads' provided all the training to the teachers through one day of training delivered in-person and undertook one subsequent visit to participating schools; further online support was also available to the participating teachers. Teachers received copies of the Teacher Handbook. Teaching materials were provided for use in the classroom. No workbooks for children were provided by the Work Group Leads to the schools, but the programme designers used Child Workbooks composed of the record sheets from the efficacy trial in the context of other EEF initiatives.

The current pilot project aims to evaluate an online training for teachers to deliver the Mathematical Reasoning programme. The course was developed by drawing on the training for the efficacy trial in content and delivery approach. As in the efficacy trial, the course content includes a description of the theoretical background (i.e. presentation of basic concepts, brief summary of supporting research), a description of the elements and key principles, and practical guidance for implementation. These elements are delivered in modules that include video lectures in a virtual learning environment, Canvas, that provides the opportunity to involve teachers in activities that aim to help them to reflect on different aspects of teaching and learning mathematical reasoning and its significance. The modules also include videos of children engaged by trained teachers in activities in the programme. The sample of videos of children engaging in mathematical activities and problems is considerably larger than in the efficacy trial and includes examples of delivery of the programme in the classroom.

Five modules that cover the aims and content of the training programme in the efficacy trial are required; completion of the first four allows teachers to access the classroom materials and the Teacher Handbook whereas the fifth module focuses on helping teachers to get ready to implement the intervention. These modules are followed by four additional modules that aim to extend the teachers' reflections about mathematical reasoning to other topics.

The Mathematical Reasoning programme that is delivered to pupils in the classroom has not changed: it has the same two strands: quantitative reasoning and arithmetic (also referred to as number sense or numerical reasoning). In order to support delivery in the classroom, the same pre-prepared PowerPoint presentations for use in the whole class activities as well as the same record sheets for children to do the work (now collected in the Child Workbook) and online games are made available to the teachers.

Appendix B: programme-level theory of change / logic model

EEF Theory of Change – Mathematical Reasoning for 6-year olds

PROBLEMS	OVERALL AIM
There is limited focus on mathematical reasoning currently in the English National Curriculum. This stems from a lack of differentiation between quantitative reasoning and arithmetic as well as from a narrow definition of arithmetic, which focuses on number facts rather than on numerical reasoning.	The overall aim is to promote an increased awareness of the need to teach children about quantitative reasoning about numerical reasoning from the beginning of primary school as essential to number sense. The Mathematical Reasoning Programme for 6-year olds focuses on these two components of mathematical reasoning. Teachers who learn to use the programme can reach a deeper understanding of these two aspects of mathematical reasoning and children who use the programme make gains in mathematical development.
TARGET POPULATION	
Year 2 teachers and pupils in English primary schools.	

	INPUTS	OUTPUTS	SHORT TERM OUTCOMES	LONG TERM OUTCOMES
TEACHERS	Resources: online CPD including interactive modules, online forum, live virtual webinars, materials for teachers	Teachers and TAs complete 9 modules of online professional development	Increased understanding of quantitative and numerical reasoning by teacher and TAs	Greater recognition of the crucial role of mathematical reasoning in primary school leading to changes in the curriculum
	Human: teacher and TA training, delivery of intervention in the classroom in regular classes over 12 lessons	Teachers receive materials for delivering the mathematical reasoning programme and obtain access to the website	Teachers share insights across the school and can have a school-wide impact on maths and science teaching	Qualification of a larger number of teacher leaders to support the online CPD in the future
PUPILS	Resources: pupils' workbooks, materials for games, access to online games	Pupils receive the 12-lesson mathematical reasoning intervention	Enhanced mathematical development as measured by standardised tests	Pupils have a better understanding of quantitative reasoning
	Human: classroom teaching with support by TA, supervision of access to online games	Pupils have practice on the use of concepts learned in the classroom	Pupils use their mathematical reasoning in other lessons	Pupils develop their number sense and numerical reasoning
	Rewards: certificates of achievement in online games, access to bonus games	Pupils enjoy learning mathematics	Pupils develop confidence and enjoyment in mathematics learning	

Appendix C: training-level theory of change / logic model