Realistic Maths Education Evaluation Protocol

Sheffield Institute of Education
Sheffield Hallam University

Education Endowment Foundation

PROJECT TITLE	Evaluation of Realistic Maths Education
DEVELOPER (INSTITUTION)	Manchester Metropolitan University
EVALUATOR (INSTITUTION)	Sheffield Hallam University
PRINCIPAL INVESTIGATOR(S)	Sean Demack
PROTOCOL AUTHOR(S)	Sean Demack, Mark Boylan, Martin Culliney & Claire Wolstenholme
TRIAL DESIGN	Two-arm, multisite four-level clustered randomised controlled trial
PUPIL AGE RANGE AND KEY STAGE	11-13 (Years 7 and 8), KS3
NUMBER OF SCHOOLS	119 (60 intervention, 59 control)
NUMBER OF CLASSES	328 (159 intervention, 169 control)
NUMBER OF PUPILS	8,142 (4,011 intervention, 4,131 control)
PRIMARY OUTCOME	GL Assessment Progress Test in Mathematics, GL PTM[1]
SECONDARY OUTCOME	Subscale(s) of GL PTM (to be published in SAP)

Protocol version history

VERSION	DATE	REASON FOR REVISION
1.0 [*original*]	May 2019	*[leave blank for the original version]*

Note/acknowledgement: The Realistic Maths Education (RME) study was proposed by the Manchester Metropolitan University (MMU) RME team and then further developed in dialogue and collaboratively with the Sheffield Hallam University (SHU) evaluation team and the Education Endowment Foundation (EEF) grant and evaluation management team. This included an intervention design and analysis workshop (Humphrey et al., 2016) that

[1] See www.gl-assessment.co.uk/products/progress-test-in-maths-ptm/

took place in February 2018. Many of the main study design elements were proposed by the RME Team. The independent randomisation, evaluation data collection and analysis of the RME study are the responsibility of the independent Evaluation Team, led by Sean Demack (Principal Investigator) and Mark Boylan (Co-Investigator). The purpose of this document - the Evaluation Protocol - is to describe that evaluation process and, in line with EEF policy, is formally authored by the Evaluation Team. However, it is informed by collaborative discussion and critical review by both the MMU and EEF teams. In particular, the background rationale and description of RME draws on previous publications by the MMU team and collaborators (these include Barmby, Dickinson, Hough, and Searle, 2011; Dickinson, Eade, Gough, and Hough, 2010; Hough, Solomon, Dickinson and Gough, 2017).

Contents

Figures

Tables

Glossary

Block - refers to RME modules taught in lessons that are consecutive or nearly consecutive.

CPD day - RME continuing professional development (CPD) day for RME intervention teachers.

Gap task - 'homework' to be completed by participating teachers between training sessions.

GL PTM - GL Progress Test in Mathematics.

MMU design school - one of a small number of schools that MMU will engage in design research activity to monitor effectiveness of CPD and project activities.

MMU RME team - the Manchester Metropolitan University team undertaking design and delivery - also shortened to 'RME team'.

Module - set of curriculum materials addressing a specific topic and designed to be taught in at least one two-week block.

Realistic Mathematics Education (RME) - approach to teaching mathematics, the focus of the intervention.

RME Intervention Teacher - a teacher who is one of two or more teachers nominated by a school to take part in the professional development prior to randomisation who teaches in a school that is allocated to the intervention condition, or is a teacher in an intervention school who joins the project to replace a teacher who leaves or is otherwise unavailable (also shortened to RME teacher).

RME Intervention School - a school with teachers and pupils in receipt of the RME intervention (shortened also to intervention school).

SAP - Statistical Analysis Plan

SHU evaluation team - the Sheffield Hallam University team undertaking the independent evaluation of the Realistic Mathematics Education trial (also shortened to 'evaluation team').

'Lesson' - each RME module is loosely divided into lessons. These are conceptual lessons, and are not intended to be treated as timed lessons.

Background

Significance

Realistic Mathematics Education and English mathematics education

Based initially on the ideas of Hans Freudenthal, Realistic Mathematics Education (RME) is a pedagogical theory developed in the Netherlands over the last 40 years. RME uses realistic contexts and a notion of progressive formalisation to help the mathematical development of pupils. It is internationally recognised, and materials based on RME are used in many countries (De Lange, 1996; Van den Heuvel-Panhuizen and Drijvers P. 2014). The approach was taken up in England in 2004 by Manchester Metropolitan University (MMU) initially with KS3 pupils (2004-7) - *Mathematics in Context* - and later with KS4 pupils (2007-10), particularly those studying towards Foundation tier GCSE Mathematics. The latter development involved, with MEI, the team producing a set of textbooks and related materials under the title *Making Sense of Maths*, published by Hodder. Both projects were independently evaluated at a later date with quantitative data on pupil outcomes of the *Mathematics in Context* analysed and qualitative data collected and analysed in relation to *Making Sense of Maths* (Searle and Barmby, 2012). In 2014-16 RME was further extended for use with post-16 pupils taking GCSE resit examinations (Hough, Solomon, Gough and Dickinson, 2017).The previous studies in England have been evaluated using a variety of mixed-method designs (Searle and Barmby, 2012; Boylan and Jay, 2017). However, it has not yet been subject to a randomised controlled trial.

RME pedagogy and materials differ from those commonly used in England in their emphasis on sustained use of realisable contexts and model building to visualise mathematical processes so that learners make sense of what they are doing. RME also differs from regular teaching in that formal mathematics develops out of students' informal mathematising, rather than the primary lesson objective being acquisition of a formal process. The process of formalisation in RME seeks to maintain a link back to the original context and model that students worked with (Van den Heuvel-Panhuizen, 2003; Hough, Solomon, Gough and Dickinson, 2017).

Mathematics education traditions in England have had a well-documented impact on classroom cultures and on student experiences and expectations. Many young people see mathematics as a question of learning rules which lead to answers based on received wisdom and the authority of the teacher (De Corte, Op 't Eynde, and Verschaffel, 2002). It can often be seen as irrelevant to everyday life and as meaningless and abstract (Boaler, 2002). The patterns of classroom interaction that are fostered by a traditional transmissionist approach to teaching mathematics can lead students to have lower expectations of themselves as well as of mathematics. It is widely acknowledged that many students become disaffected with school mathematics (Swan, 2006; Nardi and Steward, 2003; Lewis 2011). Potentially more meaningful mathematics may lead to enhanced student engagement and interest.

RME, arguably, aligns with more recent developments such as the new emphasis at GCSE on solving non-routine problems and, consequently, the implications of this for early secondary mathematics. Current interest in East Asian mathematics education under the title of 'teaching for mastery' (Boylan et al. 2018) has led to an increased interest in the use of models and representations, mathematical meaning and developing the capability of pupils to apply mathematics to non-familiar contexts (Drury, 2014). Current encouragement to spend more time on mathematical topics to deepen understanding (NCETM, 2014) directs attention to pedagogies that slow down the process of formalisation. RME potentially addresses this.

The interest in East Asian mathematics education has also led some schools to explore alternatives to grouping pupils by prior attainment (setting and/or streaming) as a way of addressing the 'long tail' of low achievement in mathematics that is more prevalent in students who are disadvantaged socioeconomically (Jerrim and Choi, 2014). It is believed that the RME approach may be particularly well suited to all types of attainment teaching contexts (Hough, Solomon, Gough and Dickinson, 2017).

Realistic Mathematics Education pedagogy

RME pedagogy is distinctive in the following ways:

Use of context

Rather than being used solely in order to illustrate the applicability and relevance of previously learned mathematics (as is customarily the case in English mathematics classrooms), context is used as a sustained underpinning for developing mathematical understanding (Fosnot and Dolk, 2002; Treffers, 1987). Contexts are carefully chosen for their potential for model building, and teachers encourage students to develop and refine their informal mathematical strategies and models. Contexts can be taken from the real world, from fiction or from an area of mathematics that students are already familiar with, but they need to be imaginable for students, and so able to support engagement with purposeful mathematical activity (Gravemeijer, Van den Heuvel-Panhuizen and Streefland, 1990).

Use of models

Models are used to bridge the gap between informal understanding and formal systems in ways that allow:

- the formal and informal to 'stay connected' in the minds of students;
- mathematical activity at differing levels of abstraction; this enables learners who find more formal concepts difficult to engage with to make progress and develop strategies for solving problems.

The meaning of 'model' extends to "materials, visual sketches, paradigmatic situations, schemes, diagrams, and even symbols" (Van den Heuvel-Panhuizen, 2003).

Mathematising

RME promotes two ways of 'mathematising' - (i) solving the contextual problem under consideration ('horizontal mathematisation'), (ii) working within the mathematical structure itself by reorganising, finding shortcuts, and recognising the wider applicability of their methods: this is called 'vertical mathematisation' (Treffers, 1987).

Multiple strategies

Both formal and informal strategies are valued and consequently, multiple strategies feature in RME lessons. The aim is for students to become more mathematically efficient and sophisticated over time; this efficiency and sophistication is not taken as a starting point but as 'on the horizon' (Fosnot and Dolk, 2002), and informal methods are still valued (Webb et al., 2008).

Redefining progress

In RME, progress is defined in terms of the progressive formalisation of models (Van den Heuvel-Panhuizen, 2003), and in particular the progression from 'model of' to 'model for' (Streefland, 1985). Initially, the model is very closely related to the specific context being considered, but eventually becomes a model which can be applied in numerous mathematical situations (Van den Heuvel-Panhuizen, 2003).

RME, classroom culture and affect towards mathematics

As a result of the above distinctive features, classroom cultures where teachers follow RME pedagogy differ from those based on more prevalent forms of mathematics teaching. RME classroom culture is underpinned by different socio-mathematical norms (Yackel & Cobb, 1996) which involve more active mathematising by pupils, and discussion and sharing of strategies. Specifically, an RME classroom culture might be marked by:

- time spent talking about context which might appear to be non-mathematical

- time spent generating/discussing various representations of a context
- systematic provision of spaces in which students are invited to talk about their strategies for solving even apparently straightforward questions
- teacher questions which are open rather than closed
- student willingness to initiate/share/discuss/question/explain strategies.

Potentially, learners will be more engaged, as mathematics is more meaningful and some features of regular teaching that are alienating diminish (see Nardi and Steward, 2003).

However, affect in mathematics classrooms is complex. It has been noted that RME may be resisted by students who have become used to particular mathematics classroom cultures: whilst students may not like their usual classroom experience, they are familiar with it and may have developed strategies that enable them to participate on some level (Hough, Solomon, Gough and Dickinson, 2017).

Evidence for the efficacy of RME

Three types of research evidence are noted in this section: international evidence on the use of Realistic Mathematics Education; evidence from studies in England; and evidence for related practices.

RME is employed by over 80% of schools in the Netherlands, which is considered one of the highest achieving countries in the world in mathematics (TIMSS 1999, 2007, 2010; PISA 2000, 2006, 2009, 2015). In PISA 2015, the Netherlands was ranked 11th in mathematics (compared to the UK in 27th place). However, in science the Netherlands was ranked 17th (behind the UK) and in reading it was ranked 15th. Thus, this indicates that national mathematical success in the Netherlands may be due to specifics of mathematics education and not only be due to the education system in general. A recent international comparison of numeracy levels amongst 16-18 year olds by OECD showed the Netherlands in 2nd position and the UK in 17th (Department for Business, Innovation and Skills, 2013).

An evaluation of the application of RME in the US - Mathematics in Context - led to a large number of detailed monographs and other outputs[2]. Comparison of intervention classes with high fidelity of implementation, with classes that followed conventional teaching, identified an association between RME and higher achievement (Romberg and Shafer, 2005). MMU's previous implementation of RME in England has been evaluated. Comparative analysis was undertaken of performance of 50 Y7 pupils who had experienced RME and 50 comparison group pupils. The pupils experiencing RME did so

[2] http://micimpact.wceruw.org/

over a year through the use of Mathematics In Context textbooks, during an initial pilot in 2004/05, (Dickinson & Eade, 2005). Details of the extent to which RME methods and materials were used by their teachers is not provided in the report. Analysis focused on a test of students' capacity to solve nine problems and explain their approaches. This found that the intervention group were both more likely to solve the problem, and to be able to explain their methods (Searle and Barmby, 2012). An independent evaluation of the post-16 GCSE project involving delivery of a number module (12 hours) and an algebra module (9 hours) found an initial positive effect on intervention students for the number module, but not of the algebra module, in comparison with students not experiencing RME. However, the positive effect was not sustained in a delayed post-test when students were retested 3 months later (Boylan and Jay, 2017).

A number of RME-type practices are advised in recent guidance for KS2 and KS3 mathematics teaching (EEF, 2017; see also Hodgen, Foster, Marks and Brown, 2018). A number of the evidence-based recommendations appear closely aligned with RME: use of representations, focusing on problem-solving strategies, developing pupils' independence; and using tasks and resources to challenge and support pupils' mathematics.

Intervention

1. Brief name

Realistic Maths Education

2. Why

The intervention aims to improve the quality of mathematics teaching in lower secondary school. The intervention provides professional development that aims to prepare teachers to confidently use RME materials, and be able to adapt other existing mathematics materials to extend the approach to their overall mathematics teaching. Table 1 in the Appendix provides indicative content of CPD days.

At the end of this section, two 'theory of change' models are presented (Table 1 and Figure 1). The first model provides a programme theory of change (Weiss, 1997) - that is it aims to model posited causal relationships between different components. This model includes some of the complex interrelationships and feedback loops that may be involved. Secondly, an implementation theory of change model is provided in a format that summarises inputs and expected outputs.

3. What (materials)

The intervention curriculum materials are organised in modules[3]. Modules are loosely divided into lessons which comprise sets of activities. As well as providing printed copies of materials at CPD events, participants are also able to access material on a section of the project website.

Teachers are also provided with teacher guides and supporting materials (PowerPoints, worksheets) which will enable them to teach 5 two-week blocks of RME in each of Years 7 and 8.

4. What (procedures)

Professional development:

Teachers involved in the intervention receive specialist training in how to use RME materials and develop their practice over the two years of the trial. Professional development focuses on both effective use of the curriculum materials, and on developing the underlying RME pedagogical principles. Professional development combines off-site CPD days and in-school tasks. The intervention can be considered a 'curriculum professional learning' innovation (Boylan and Demack, 2018) in that the CPD itself is intended to lead to pedagogical change, and so to pupil impact, but that also the use of the curriculum materials has a potential professional learning effect, as well as directly impacting on pupil learning.

RME teaching:

Curriculum materials are organised into materials for Y7 and then Y8, with the following modules in each year: number, proportional reasoning, data, geometry, and algebra. Each lesson or set of activities (outlined above) may last more than one teaching lesson depending on lesson timing. Teachers work through the activities, deciding for themselves if they will omit any. Each module consists of approximately 8 lessons' worth of materials. As the approach promotes flexibility, a lesson might last 1 hour, but could be extended to 2 hours or more. One lesson will involve a mean of approximately 3 activities. Overall, the module design involves giving teachers more material than they might need for two weeks, and teachers are encouraged to take more time if they feel that this is appropriate.

5. Who (provider)

RME curriculum materials are based on materials previously developed by MMU, updated and refined for current teaching contexts

[3] As indicated by Table 2 in the Appendix

6. How

Professional development components are:

- An eight day course: teachers will attend 8 training days between October 2018 and April 2020; the course will involve demonstration lessons, training in RME principles (e.g. curriculum design and pedagogy) and discussion of teachers' experience in 'gap tasks'.
- Between sessions, teachers will work with the RME materials and work collaboratively to evaluate their practice and undertake gap tasks.
- Use of RME materials and scheme of work guidance for Years 7 and 8 (see below - curriculum materials).
- Access to online resources: teachers will have access to a dedicated website in order to support them in gap tasks and the use of materials, all available in electronic form.

A key component of the professional development is gap tasks which encourage teachers to reflect on their RME practice with their paired teacher. Gap tasks are to be done between CPD sessions, and focus on developing diagnostic insights and pedagogic practices which support an RME classroom culture. The aims and purposes of gap tasks are to:

- encourage focus on teachers' pedagogic practice through 'noticing' and so develop insight into:
    - how much time is spent using particular activities and resources
    - 'wait time' and related issues
    - response to student strategies and contributions
    - orchestrating and responding to disagreement/mathematical argument
    - physical positioning in the classroom etc.
- develop diagnostic skills in analysing student work:
    - understanding student approaches and strategies
    - moving away from deficit views of student work
    - understanding how materials do or do not support development
- practice adapting existing materials towards an RME approach
    - use of local contexts/interests
    - thoughtful selection of materials and activities
    - reflection on how materials progress through ideas
    - anticipating how students respond to tasks

Gap tasks will frequently involve observation, but it is recognised that teachers may often fail to find time to observe each other. Although this will be encouraged, teachers will also be asked to video themselves and reflect alone and with their paired teacher on particular pedagogic strategies. Gap tasks will focus on specific aspects of RME pedagogy, and encourage teachers to think about this in their lesson planning and work on this aspect in their lessons, and in lesson observation with their paired teacher. This practice will draw on and aim to develop 'the discipline of noticing' (Mason, 2002) in terms of teachers' thinking about how students' progress is supported in a lesson, and how their practice enables this. The gap tasks are designed to enable teachers to focus on one particular strategy at a time, in order to build towards RME 'fluency'.

7. Where

CPD will take place in 6 geographical hub areas and teachers will then work within their own schools in teaching RME.

8. When and how much

CPD days will take place between Autumn 2018 and Spring 2020, provisionally in Oct, Dec 2018, Feb, Apr, Jun, Oct 2019, and Feb and Apr 2020.

Teaching will take place over two school years between autumn 2018 and summer 2020.

9. Tailoring

In addition to refinement of materials specifically for this intervention, the CPD framework includes one 'catch up' CPD day with incoming Y8 teachers to support continuity of RME experience for pupils whose Y7 teachers do not continue with the programme.

10. Modifications

None at present.

Theory of Change

Causal mechanisms

The intervention involves a number of potential change agents and related mechanisms or processes that may lead to change either directly or indirectly in pupil outcomes; these are described in table 1 below:

Table 1 Theory of change mechanisms

Causal agent	Mechanism/process	Outcome/desired change (intermediate or end point)
RME professional development events	Indirect: Professional learning	Teacher: Change in any or all of knowledge, belief, practice - teaching and collegial, values
In school PD tasks	Indirect: Professional learning	Teacher: Change in any or all of knowledge, belief, practice - teaching and collegial, values
Use of RME materials	Indirect: 1. Professional learning	Teacher: Change in any or all of knowledge, belief, practice - teaching and collegial, values, culture
	Indirect 2: Pupil and class change	Class: Foster RME culture (use of context, expanded notion of progress/success, mathematising, use of models/representations, use of multiple strategies) Pupil: capacity to use context, expanded notion of progress/success, mathematising, use of models/representations, use of multiple strategies
	Direct: Pupil engagement with RME maths	Pupil mathematical learning and capacity to use context, expanded notion of progress/success, mathematising, use of models/representations, use of multiple strategies
Use of RME principles in other mathematics lessons	Indirect: Foster an RME culture Direct:	Pupil mathematical learning and capacity to use

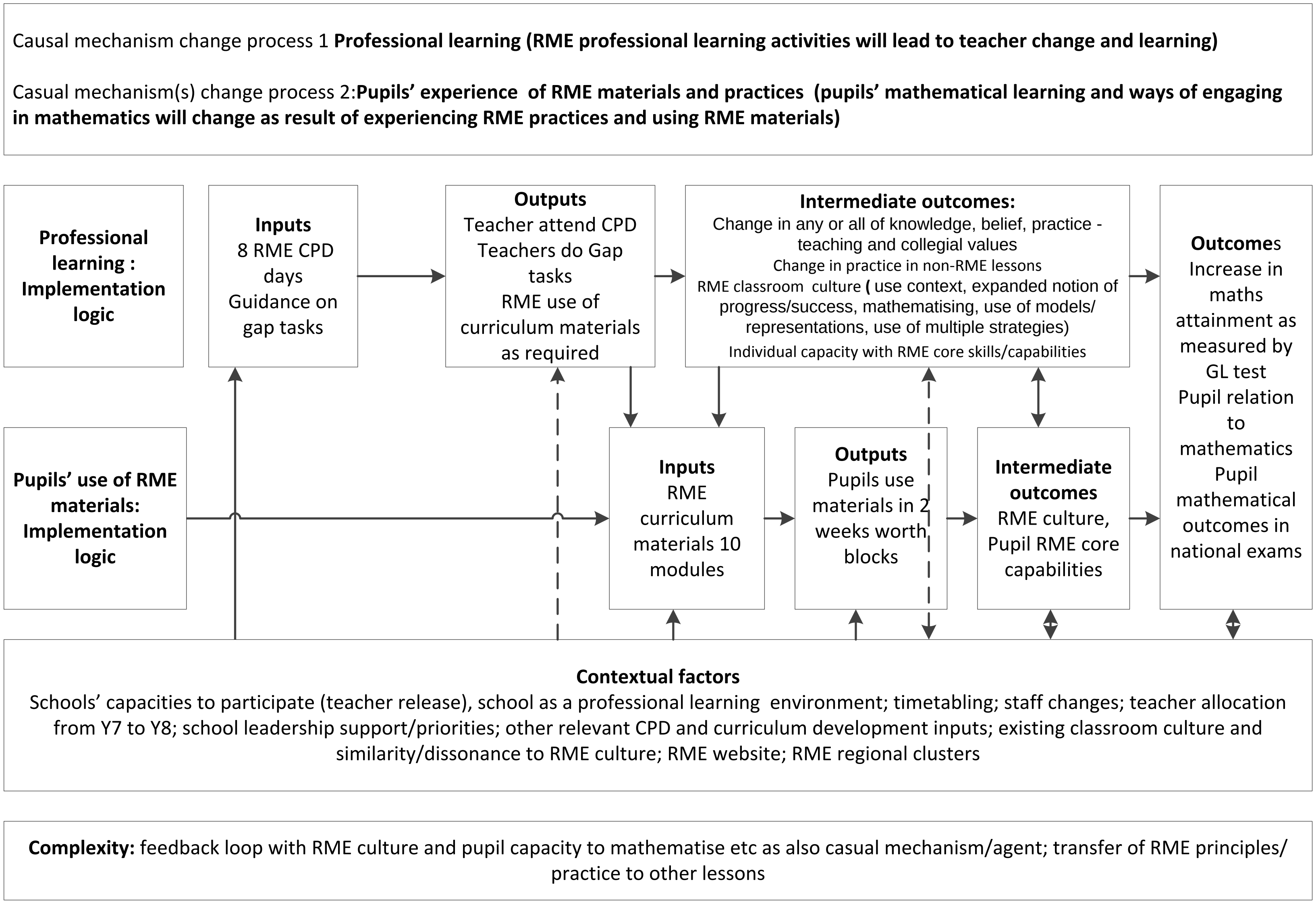
	Enhanced pupil mathematical activity	context, expanded notion of progress/success, mathematising, use of models/representations, use of multiple strategies
The development of an RME culture (use of context, expanded notion of progress/success, mathematising, use of models/representations, use of multiple strategies)	Direct: Enhanced pupil mathematical activity	Pupil mathematical learning and capacity to use context, expanded notion of progress/success, mathematising, use of models/representations, use of multiple strategies

It can be seen that the latter two mechanisms/processes are both posited as intermediate outcomes and causal processes.

Logic model

Figure 1 below provides a logic model that links main inputs and outputs. For simplicity RME culture is presented as an intermediate outcome only, as is the use of RME principles in other lessons. Two potential causal processes are foregrounded in this logic model: RME professional development as an indirect change process on pupil outcomes and the use of RME materials as a direct process.

Figure 1 Logic model



Research Plan

Research questions

The evaluation aims to address the following questions:

Impact evaluation

Primary question

RQ1. Does the RME intervention improve pupil attainment in mathematics over 2 years as measured by the GL Assessment Progress Test in Mathematics (PTM) for the intention to treat group compared to the control group, in general and specifically for FSM pupils?

Secondary questions

RQ2. What is the impact of RME on mathematics attainment for pupils known to have been taught by one of the nominated RME teachers throughout the trial period?

RQ3. What is the impact of RME on mathematics attainment for pupils who experience partial intervention effects due to pupil and/or teacher movement during the trial period?

RQ4. What is the effect on attainment of components of the GL PTM that most closely related to the RME curriculum material content and on questions identified as related to problem solving?

RQ5. What is the relationship between mathematics attainment and fidelity of implementation?

RQ6. What is the impact of RME on mathematics attainment for pupils taught by teachers identified as implementing the RME evaluation with high fidelity?

Implementation and process evaluation questions

RQ7. What is the fidelity of participation in CPD and implementation of RME school-based gap tasks, and use of materials by nominated teachers?

RQ8. What has influenced variation in implementation by nominated teachers?

RQ9. What are teachers' perceptions of differences between RME practices and those they used prior to the intervention; and what are the differences in reports of practices between teachers participating in RME and those in control schools?

RQ10. How do teachers change as a result of participation in the intervention, and in particular in relation to beliefs about mathematics and teaching mathematics, and self reported changes (if any) in mathematics pedagogical knowledge, pedagogical content knowledge, and mathematics content knowledge?

RQ11. What issues are important, if any, for the security of the trial outcome, replication and scalability?

In addition to the evaluation, the MMU RME team will be undertaking their own, internal design research in a number of 'design schools' (see below for details on IPE samples). Foci for this activity are: teachers' engagement with RME pedagogic practices, teacher use of lesson time and materials, the relationship between teacher practice and professional development experience, including how teachers engage in 'gap tasks', the development of RME 'classroom culture', and analysis of student work in relation to RME approaches and models. MMU's design school research questions are provided in Appendix II of the protocol.

Design

The two year intervention (2018-2020) is both a curriculum and professional development project. 119 schools were recruited to take part in the trial (60 allocated to the intervention and 59 to the control condition). In each intervention school, at least two teachers will participate in professional development and then teach curriculum materials to Y7 pupils starting from October 2018. Ideally, the same two teachers will then teach the pupils in Y8, until the end of 2019/2020, using materials designed for Year 8 pupils, and will continue with CPD. However, where the original nominated teachers are unavailable (for reasons of staff turnover, other absence or other staffing requirements), schools will nominate replacement teachers who will teach the same pupils who experienced RME in Y7, and who will attend the training programme, to provide continuity. Whilst the intervention is designed for two teachers per school, in order to support in-school collaborative professional development opportunities, schools can nominate more than two teachers to take part in CPD in each year if they so choose.

Pupil recipients will be primarily the pupils of the two teachers nominated prior to randomisation, thus will be approximately 60 per school. The primary Intention to Treat (ITT) impact analysis (see below) will focus only on pupils who were located in one of the Y7 maths classes that were taught by one of the two RME teachers nominated prior to randomisation. However, it is anticipated that due to movement between classes over the two years, some pupils will have a partial experience of RME. Thus, all Y8 pupils in the intervention schools will be tested in 2020 and this data will be included within the evaluation impact analyses.

Table 2 Trial design

Trial type and number of arms		Two-armed multisite clustered randomised controlled trial
Unit of randomisation		School
Stratification variables (if applicable)		Geographical Area (6 areas); School phase (secondary / middle) & use of setting/streaming in Y7 maths (yes/no)
Primary outcome	variable	Maths attainment
	measure (instrument, scale)	GL Progress Test in Mathematics (PTM)
Secondary outcome(s)	variable(s)	Subscales of maths attainment
	measure(s) (instrument, scale)	PTM subscales (to be finalised & published in SAP)

The impact of RME on mathematics attainment will be evaluated using a two-arm multisite clustered randomised control trial (MSCRT) research design with randomisation at the school level. This design is what Spybrook et al (2016) classifies as a 4-level MSCRT but the eventual multilevel model will only include three levels (school, class and pupil). Schools will be randomised within each of the six geographical areas/sites[4]. Because this is an efficacy trial, the design has not been powered to detect variation in effect sizes across geographical areas. The fourth (multisite) level is accounted for in the design via the degrees of freedom for the impact evaluation power analyses (see below, p23). Geographical area will not be included as a level in the model, but will be included through introducing dummy covariates in the 3-level regression model (see below).

Randomisation will take place at the school level for practical and methodological reasons. Practically, for recruitment and implementation of the RME intervention, a school level approach is preferable. Methodologically, school level randomisation reduces the risk that the RME intervention might 'spill over' from the intervention to the control group. For example, in a within-school class level randomisation the likelihood of intervention teachers and/or pupils sharing RME materials and/or experiences with control teachers or pupils is higher than a school level randomised design. School-level randomisation does not eliminate the possibility of spill over (e.g. teachers/materials moving between schools within a Multi-Academy Trust), yet it does provide stronger protection than within-school randomisation.

Prior to school recruitment the randomisation planned to draw on KS4 school census data in order to undertake a propensity-score-paired-school-stratification (Boylan et al., 2018).

[4] Six areas: NE & Yorkshire; NW; East Midlands; West Midlands; South; London

However, in order to maximise school recruitment numbers within a fairly tight recruitment period (Jan to July 2018, with schools signing MOUs from May), two pragmatic decisions made this approach unfeasible. First, the RME programme was offered to middle schools which have no KS4 statistics because pupils move on from middle schools to secondary prior to taking KS4. Second, RME was offered to secondary schools that were 'new' or had recently changed governance structure (e.g. becoming an academy). In some cases where governance had changed, KS4 statistics were available for the previous incarnation of the school but this was not the case for all schools. Therefore, a simpler approach to randomisation was adopted that did not rely on school level KS4 statistics. A stratified randomisation approach was adopted that included controls for geographical hub region, whether a school was a middle or standard secondary school and whether a school currently had a policy of using within-school selection (setting/streaming) to sort pupils into maths classes in Y7 or not. Using these three controls best ensures that the RME intervention and control samples are balanced across the geographical hub regions, have similar numbers of middle and secondary schools and schools that have setting/streaming within-school selection policies.

Prior to randomisation, schools that signed the RME MoU were required to submit two sets of data. First, between June and September 2018, schools were required to specify whether setting/streaming policies were used in Y7 maths and the names of at least two nominated teachers to participate in the RME programme if their school was randomly selected to the intervention group. Second, in September 2018, schools were required to submit pupil lists for all Y7 maths classes attached to named teachers. Once both sets of data were submitted and checked to confirm that nominated teachers were clearly attached to a specific Y7 maths class, a school had completed all that was needed prior to randomisation. As part of the IPE, a baseline teacher survey was also undertaken in September 2018. Completion of this survey was voluntary and so not a requirement for randomisation but the survey was closed prior to randomisation[5] to ensure that teacher responses were not influenced by their school's allocation.

A total of 129 schools were recruited to the RME evaluation, 119 schools provided pupil data to the evaluation team and were therefore entered into the randomisation process. Randomisation took place on Friday 5th October 2018. 110 schools had provided the necessary data and were informed of their allocation immediately after randomisation. The allocation for the remaining nine schools was withheld until we had received and checked the two sets of required data. Eight schools were informed of their allocation prior to the first week of RME CPD training and one school was informed between the first and second training

[5] The RME delivery partners at MMU requested that the survey be re-opened following this in order to maximise the baseline response. We obliged this request but, for the evaluation, will only be using data collected prior to randomisation.

days.

Participants

Recruitment for this trial was overseen by the MMU RME Team. To keep variation at the school level to a minimum, the aim was to recruit 120 non-selective established secondary schools for the evaluation. As noted above in the design section, this criterion was relaxed at towards the end of June to expand the offer to middle schools and to new schools in order to maximise the chance of reaching the target of 120 schools.

Schools were required to commit at least two teachers to the RME CPD training to take part. These teachers should be timetabled to teach all of their Y7 mathematics classes each week; however classes split between two teachers involved in the intervention could be considered. Whilst all Y8 pupils will be tested, and this data will be used within the wider impact analyses (see below), the primary Intention to Treat (ITT) impact analysis will focus just on pupils who were located in the Y7 maths classes that were taught by one of the two RME teachers nominated (or more than two if a school commits additional teachers) at the start of the trial in September 2018. It was assumed that the nominated RME teachers will teach an average of three Y7/Y8 maths classes between them. With a mean of 25 pupils per class, the ITT analysis will involve a mean of 68 pupils per school and a total of 8,142 pupils across the 119 schools.

Outcome Measures

The primary outcome measure is maths attainment, to be assessed through the GL Progress Test in Mathematics (PTM), administered by the evaluators using independent invigilators who will be blind to whether a school is in the intervention or control group. The GL PTM will be taken by all Y8 pupils in June 2020. The GL PTM is a commercial standardised assessment of pupil mathematical skills and knowledge[6]. GL PTM was selected as the primary outcome for four reasons. First, it is well known and widely used by schools and teachers. Second, it is aligned to the mathematics national curriculum. Third, it enables the assessment of pupil mathematics at the end of Y8. Finally, it enables the assessment of mathematics attainment within curriculum areas and mathematical capabilities. The fourth reason relates to the creation of secondary outcomes for the impact evaluation in order to address RQ4. During the evaluation, PTM curriculum areas / mathematical capabilities will be selected that are closely related to the RME curriculum and on 'problem solving' and this will be published in the Statistical Analysis Plan which will be finalised by Spring 2019.

[6] www.gl-assessment.co.uk/products/progress-test-in-maths-ptm/

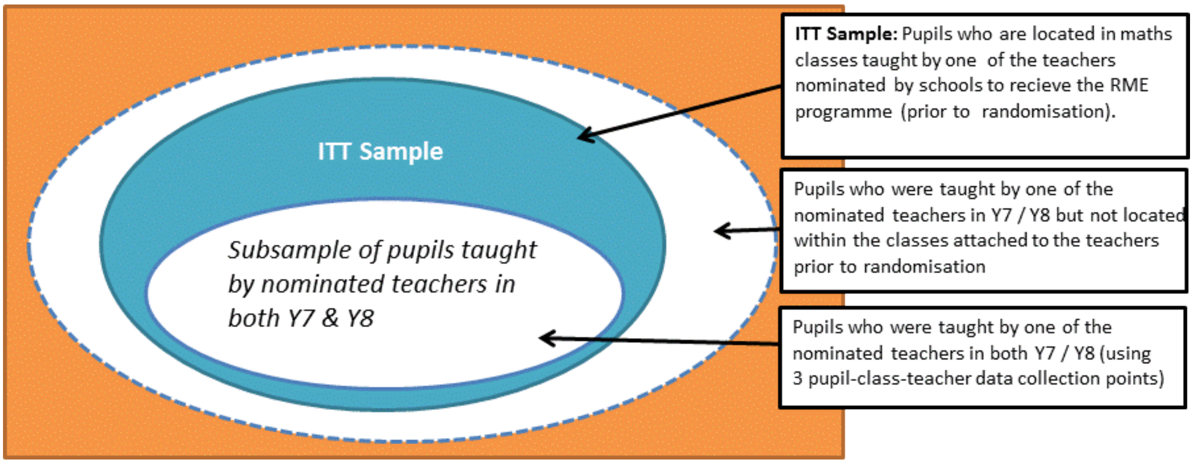
Table 3 Sample size calculations

		OVERALL	FSM
MDES			
Pre-test/ post-test correlations	level 1 (pupil)	0.70	0.70
	level 2 (class)	0.70	0.70
	level 3 (school)	0.70	0.70
Intracluster correlations (ICCs)	level 2 (class)	0.05 to 0.50 (see below)	0.05 to 0.50 (see below)
	level 3 (school)	0.15-0.20	0.15-0.20
Alpha		0.05	0.05
Power		0.80	0.80
One-sided or two-sided?		2	2
Average classes per school		3	3
Average pupils per class		25	5
Number of schools	Intervention	60	60
	Control	59	59
	Total	119	119
Number of classes	Intervention	159	159
	Control	169	169
	Total	328	328
Number of pupils	Intervention	4,011	900
	Control	4,131	885
	Total	8,142	1,785

For the power analyses, we assume a statistical significance of 0.05 or less, and a statistical power of 0.80 or higher. We also assume that there will be 20 schools in each of the six cluster areas. For the ITT analyses, we assumed three Y8 maths classes per school, and 25 pupils per maths class. From the baseline data collection, the mean number of classes per school was 2.8, and the mean pupils per class was 24.8.

Figure 2 summarises the Y7/Y8 pupil population for the 119 recruited schools and three key pupil samples. The Intention-To-Treat (ITT) sample shown in blue will include Y8 pupils who were identified as being located in a classroom taught by one of the two or more maths teachers nominated by schools prior to randomisation. This will be the sample used for the primary ITT impact analyses (RQ1).

Figure 2 Pupil Samples and Subsamples for the RME impact evaluation



Class lists of Y7 and Y8 maths classes will be monitored through the trial period. This monitoring data will be drawn on to identify specific pupil samples for follow-on impact analyses (to answer RQ2 and RQ3). Restricting the ITT subsample to only include pupils who were taught by a nominated teacher in both Y7 and Y8 (RQ2) ensures that the sample will have had consistent exposure to RME in their maths classroom for both years of the trial. Finally, a pupil subsample known to have been taught by one of the nominated teachers for at least some of Y7/Y8 can be identified (RQ3), enabling assessment of whether partial exposure to RME led to a positive impact on maths attainment.

The strength of class-level clustering of the GL PTM outcome will depend on the use of setting/mixed ability teaching across the 119 recruited schools. For the school and class level ICC estimates used in the power analyses, we draw on our evaluation of the Multiplicative Reasoning Project (Boylan et al., 2015, henceforth MRP). Power analyses are structured around three scenarios relating to the use of setting / mixed ability teaching in Y7/Y8 maths across the sample of schools.

Scenario 1: All 119 school practice mixed ability in all Y7 / Y8 maths classes. School level ICC =0.15-0.20; **class level ICC=0.05**.

Scenario 2: Some of the 119 school practice mixed ability in all Y7 / Y8 maths classes others use setting. School level ICC =0.15-0.20; **class level ICC=0.25.**

Scenario 3: All of the 119 school practice setting in all Y7 / Y8 maths classes. School level ICC =0.15-0.20; **class level ICC=0.50**.

Within the MRP evaluation, we found that including a baseline covariate at the pupil level resulted in statistically accounting for variance at pupil, class and school levels. To maximise statistical precision, we will use KS2 mathematics attainment (taken in summer 2018) as a baseline covariate. A correlation of 0.70 will be adopted as an estimate between KS2 attainment and the GL PTM outcome (Allen et al., 2018). The power calculations are informed by the Optimal Design software, however (as shown in Appendix I) we have used the formulae from Spybrook et al (2016), so that explanatory power can be included at pupil, class and/or school levels and avoid limitations of Optimal Design estimates[7]. The same explanatory power (R^2=0.49) at all three levels for the formula-based MDES estimates is adopted.

Because this is an efficacy trial, we assume that the effect size variability across the six geographical clusters will be zero. This means that geographical area is acknowledged in the design through the calculation of degrees of freedom and in the analyses through the inclusion of dummy covariates in the 3-level (school, class & pupil) model.

Table 4 summarises our MDES estimates for the three scenarios in terms of grouping by prior attainment. In England the most common term for alternatives to setting and streaming (grouping by perceived ability) is 'mixed ability' however, we use the term heterogeneous grouping, on the understanding that this is a relative term. This is done for the main ITT analysis (3 classes, 25 pupils per class), for a subgroup ITT analysis amongst pupils classed as FSM (3 classes, 5 FSM pupils per class) and for an analysis that includes the whole of Y8 (6 classes, 25 pupils per class).

- 4 level MSCRT (pupils > Y7/Y8 maths classes > schools; blocked by geography)
- $p<0.05$; statistical power 0.80; Number of sites = 6; Schools per site = 20
- Pupil level baseline covariate (KS2 maths) that accounts for 49% of the variance at all levels (school, class and pupils) for formula estimates and just at level 3 (school) for OD estimates.
- Randomisation at the school level; A balanced design (60 intervention schools; 60 control schools)
- School level ICC = 0.15-0.20
- Class level ICC allowed to vary between 0.05 (scenario 1) and 0.50 (scenario 3)

[7] Optimal Design only allows covariates to be included at the school level within a 4 level MSCRT.

Table 4 MDES estimates for RME evaluation

	Class Level ICC	MDES Estimates	
		Formula	OD Software
Main ITT analysis: **teachers/classes/pupils identified prior to randomisation (3 classes per school, 25 pupils per class)**			
1 (all heterogeneous)	0.05	0.16-0.18	0.17-0.19
2 (some heterogeneous)	0.25	0.18-0.20	0.21-0.23
3 (all setting/streaming)	0.50	0.21-0.23	0.26-0.27
Subsample Analyses: ***teachers/classes identified prior to randomisation - pupils ever classed as FSM (3 classes per school*; 5 FSM pupils per class[8])**			
1 (all heterogeneous)	0.05	0.17-0.19	0.20-0.21
2 (some heterogeneous)	0.25	0.19-0.21	0.23-0.24
3 (all setting)	0.50	0.22-0.23	0.27-0.28

For the main ITT analysis, a 3 level MSCRT research design that ignored the class level[9], results in an MDES estimate of 0.18 standard deviations, which echo the 4 level MSCRT estimate when all recruited schools are assumed to be practicing mixed ability teaching (i.e. when class level clustering is minimised). Given that this is not a reasonable assumption for Y7/Y8 maths teaching in England (Francis et al., 2017), ignoring class level clustering is likely to result in a trial design that is either underpowered and/or overstates statistical precision.

Our MDES estimates are dependent on the use of mixed ability / setting within Y7/Y8 across the 119 recruited schools. From the initial school data collected June-Sept 2018, 106 of the 119 (89%) recruited schools reported to use some form of setting or streaming in Y7 maths. The specific approach to setting/streaming varied across the 106 schools with some schools reporting that all Y7 pupils are placed into maths classes based on KS2 maths attainment and others reporting to use a 'diamond' structure with a top and bottom set with mixed ability groups in-between. This suggests that the MDES estimates relating to scenario 2 or 3 will be the more accurate for this trial. KS2 maths attainment data (from June 2018) will be obtained from the NPD to provide more precise school and class level clustering estimates for an updated power analysis. It should be noted that ICC estimates based on KS2 maths may not reflect those found with the PTM outcome. Amongst other things this might be because over the course of the two year trial, schools introduce (or move away from) setting within Y7 / Y8 mathematics. However, the KS2 ICC estimates will provide an empirical estimate on the extent of class-level clustering at baseline and this will be used to update the MDES estimates

[8] Estimated as 14% of secondary pupils from https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2017
[9] Specifically, this is a 2-level CRT (pupils clustered into schools, ignoring class level clustering) blocked by geographical area but assuming zero effect size variation between areas

for the trial which, assuming we obtain this NPD data in time, will be published either in an updated protocol or the Statistical Analysis Plan.

Analysis plan

Intention to treat analyses of the primary outcome (RQ1)

Table 5 below summarises the planned main impact analyses for the primary outcome of the RME evaluation.

A multilevel approach will be taken, with pupils clustered into classes and classes clustered into schools. Multilevel linear regression models will be constructed for the GL PTM primary outcome. KS2 maths attainment will be used as the baseline covariate. The first model will only include the school level group identifier (an outcome only model). The second model will also include KS2 maths attainment as a covariate at the pupil, class and school level[10]. The final model will also include the three variables[11] used within the stratified randomisation (geographical hub area, setting/streaming dummy and middle/secondary dummy). It will be this final model that will be used for the headline ITT impact analysis for the PTM primary outcome.

Follow-on ITT analyses will focus on the impact of the RME programme on maths attainment for disadvantaged pupils[12]. The same three model stages used for the headline ITT analyses will be used here.

For each model, the coefficient of the school-level dummy variable used to distinguish 'intervention group' pupils within the 60 schools who will receive the RME programme from 'control group' pupils will be converted into Hedges' *g* effect size statistics with 95% confidence intervals.

[10] These will be centred so that the school level will be centred on the mean for all 119 schools; the class level will be centred around the school mean; the pupil level will be centred around the class mean.

[11] This is actually seven binary variables; five will be used for the six geographical areas, one for school phase (secondary / middle) and one for use of setting/streaming (yes/no)

[12] The NPD KS4_FSM6CLA1A variable, pupils who have ever been "eligible for free school meals or has been looked after for a day or more or has been adopted from care".

Table 5 Summary of analysis plan for ITT impact analyses for primary outcome (RQ1).

Multilevel linear regression with three levels (pupils clustered into classes clustered into schools)

Primary outcome: GL Progress Test in Maths (PTM) taken in June 2020.

Analysis and Sample	*Level 1 (pupil) Variables*	*Level 2 (class) Variables*	*Level 3 (school) Variables*
Main Analysis, ITT sample	-	-	Dummy (1=RME; 0=control)
Main Analysis, ITT sample	KS2 Maths	Mean KS2 Maths	Dummy (1=RME; 0=control); Mean KS2 Maths
Main Analysis, ITT sample	KS2 Maths	Mean KS2 Maths	Dummy (1=RME; 0=control); Mean KS2 Maths Geographical hub area Setting/streaming dummy Secondary/Middle dummy

Follow on analyses for the primary outcome (RQ2 and RQ3)

Table 6 below summarises further planned follow-on impact analyses for the *RME* evaluation. These analyses draw on GL PTM test data for the whole of Y8 and data that monitor pupil/teacher movement between classes during the trial period to identify two subsamples in order to address RQs 2 and 3.

RQ2 focuses on a potential 'epicentre' of impact for RME (illustrated as the central white ellipse in Figure 2). These analyses will be limited to a subsample of pupils who were located in a class known to be taught by one of the nominated RME teachers on three occasions through the Y7/Y8 trial period. These pupils would therefore be known to have had consistent exposure to a teacher involved in the RME programme in both 2018/19 and 2019/20. The analyses will examine whether this results in a concentrating effect for the impact of RME on mathematics attainment.

RQ3 focuses on the impact of RME in Y8 more widely (illustrated as the largest white ellipse in Figure 2). The analyses will be limited to a subsample of pupils who were located in a class taught by one of the nominated teachers on at least one of the three occasions through the Y7/Y8 trial period. The analyses will examine whether partial exposure to RME results in diluting the impact of RME on mathematics attainment.

Table 6 Summary of analysis plan for follow-on impact analyses for primary outcome (RQ2 and RQ3)

Impact of RME within three pupil subsamples

Multilevel linear regression with three levels (pupils clustered into classes clustered into schools)

Primary outcome: GL Progress Test in Maths (PTM) taken in June 2020.

Analysis and Sample	***Level 1 (pupil)***	***Level 2 (class)***	***Level 3 (school)***
ITT subsample (taught in both Y7 and Y8 - **RQ2**)	KS2 Maths	-	Dummy
	KS2 Maths	Mean KS2 Maths	Dummy ; Mean KS2 Maths Geographical hub area Setting/streaming dummy Secondary/Middle dummy
Y8 Subsample (Taught at some point in either Y7 and Y8 - **RQ3**)	KS2 Maths	-	Dummy (1=RME; 0=control)
	KS2 Maths	Mean KS2 Maths	Dummy ; Mean KS2 Maths Geographical hub area Setting/streaming dummy Secondary/Middle dummy

Impact on attainment in components of the GL PTM outcome (Secondary Outcome, RQ4)

RQ4 focuses on examining the impact of RME on specific aspects of mathematics attainment. Currently, GL classify PTM questions according to two sets of categories; curriculum areas and test questions:

Curriculum areas:

- Number
- Shape, Space and Measures
- Algebra
- Data Handling

The test questions can also be grouped into one of four areas of mathematical capabilities:

- Knowing Facts and Procedures
- Using Concepts
- Solving Routine Problems
- Reasoning

The approach to using these categories to identify components of the GL PTM most closely aligned to the RME curriculum and problem solving will be determined during the early stages of the trial and published in the Statistical Analysis Plan. The analytical approach adopted will be similar to the ITT analyses of the primary outcome presented in Table 5 above. We do not plan to conduct the follow-on analyses for these secondary GL PTM subscale outcomes.

Fidelity and CACE analyses (Primary Outcome, RQ5 and RQ6)

Table 7 below summarises further planned follow-on fidelity (RQ5) and CACE (RQ6) analyses for the RME evaluation.

RQ5 focuses on the relationship between fidelity to RME and pupil outcomes - it is closely related to outcomes from the IPE and analyses to address RQ8. Analyses that examine the impact of fidelity to RME on maths attainment will be limited to the subsample of pupils who were located in a class taught by one of the nominated teachers on three occasions through the Y7/Y8 trial period and will only involve teachers in the 60 RME intervention schools. Drawing on IPE data, dimensions of fidelity to RME[13] will be included as explanatory variables in order to explore their relative impact on mathematics attainment. Each dimension will be entered separately and then together as a block. Finally, as set out in Table 7 below, the dimensions of fidelity will be reduced to a single binary dummy variable that identifies RME teachers known to have 'complied'[14] to the RME programme (the CACE dummy variable).

RQ6 focuses on what is known as the Compliers Average Causal Effect (CACE) of the RME programme. For these analyses, the intervention group sample will be restricted to pupils located in a class taught by one of the nominated teachers on three occasions through the Y7/Y8 trial period. The CACE dummy variable will then be used to further restrict the RME intervention sample to pupils located in a class taught by a teacher known to have complied with RME through the Y7/Y8 trial. The 3-level design enables the CACE variable to be included at the class (rather than school) level. This avoids having to aggregate teacher-level compliance to the school level and enables the model to take account of within-school (class/teacher level) variation in compliance. Two comparison groups will be used for the CACE analysis; first, the complete ITT control group sample and second, the ITT control sample restricted to pupils located in a class taught by one of the nominated teachers on three occasions through the Y7/Y8 trial period.

[13] To be reviewed during evaluation and finalised/published in the Statistical Analysis Plan.

[14] This means teachers who engaged well with RME and closely followed the intended programme (as set out across the fidelity dimensions) - the term 'complied' is a common technical term more at home in economic/clinical trials that links to the follow-on Compliers-Average-Causal-Effects (CACE) analyses.

Table 7 Summary of analysis plan for fidelity and CACE analyses for primary outcome (RQ5 and RQ6)

Multilevel linear regression with three levels (pupils clustered into classes clustered into schools)

Primary outcome: GL Progress Test in Maths (PTM) taken in June 2020.

Analysis and Sample	*Level 1 (pupil)*	*Level 2 (class/teacher)*	*Level 3 (school)*
Fidelity Analyses (RQ5):			
ITT subsample (Intervention schools only) (taught in both Y7 and Y8)	KS2 Maths	Mean KS2 Maths Teacher level dimensions of fidelity*	Mean KS2 Maths School level measures of fidelity
ITT subsample (Intervention schools only) (taught in both Y7 and Y8)	KS2 Maths	Mean KS2 Maths CACE dummy	Mean KS2 Maths
CACE analyses (RQ6)			
ITT subsample (Intervention) (taught in both Y7 and Y8) Intervention school sample restricted further; CACE dummy=1. Complete / Raw control group	KS2 Maths	Mean KS2 Maths	Dummy (1=RME; 0=Control) Mean KS2 Maths
ITT subsample (Intervention and control) (taught in both Y7 and Y8) Intervention school sample restricted further; CACE dummy=1.	KS2 Maths	Mean KS2 Maths	Dummy (1=RME; 0=Control) Mean KS2 Maths

* The specific dimensions of fidelity and how these are measured will draw on data from the IPE and be reviewed and agreed during the evaluation. A binary CACE variable will be derived drawing on these fidelity dimensions. Details on the final dimensions of fidelity and CACE variable will be published in the SAP. See Fidelity section below.

RME fidelity

The table below provides initial fidelity criteria to be reviewed during the evaluation. In summary the components are:

Teacher level:

- RME PD attendance
- Attendance at RME PD Number Module
- Undertaking PD gap tasks
- RME Curriculum Time
- Whether RME is taught in blocks

- RME module coverage
- RME number modules coverage

School level:

- Paired Attendance
- Collaboration on gap tasks

The table below provides initial indicated criteria for the fidelity components, for teachers who progress from Y7 into Y8 implementing RME. These will be subject to review when the statistical analysis plan is written. In particular the following will be specified:

- what amount of material use or activity use constitutes having taught the modules
- for each gap task what constitutes 'completion'
- for each gap task what constitutes adequate 'collaboration'
- what patterns of teaching constitute teaching in a block

To derive overall fidelity - minimum or high, each separate criterion will need to be met at the specified level. The definition of what constitutes a minimum or high level of fidelity will be agreed between the evaluators and delivery partners during the trial period and published in the SAP.

For teachers who join the project at the end of Y7, completion of the catch up CPD will be required. If the catch up CPD is completed, then any recorded implementation of the teacher who taught their Y8 class in Y7 will be transferred over to them.

Implementation & Process Evaluation

Outline and aims

The IPE approach blends quantitative and qualitative methods (Humphrey et al. 2016) and is informed by both the rationale and description of the RME intervention (see above) and the theory of change logic model. In order to understand and gain insight into change mechanisms a series of in-depth case studies will be conducted (Stake, 2013).

The aim of the IPE is to provide evidence that supports the interpretation of impact findings, including providing insights into possibly causal explanations. The IPE focuses on:

- fidelity (RQ7),
- implementation including variation in this (RQ7, 8),
- teacher perceptions of differences between RME and business as usual (RQ9)

In addition the theory of change model will be interrogated by considering:

- effects on teachers' and pupils' activity (RQ10)

Finally, implications for scalability and replications will also be addressed (RQ11)

Samples

The following samples of teachers and pupils will participate in IPE data collection with details of how these samples will be determined:

Intervention

1. *All RME schools: 60 schools*
2. *All RME teachers: 120 teachers*[15]
3. *RME teachers interview sample: 7 teachers*
4. *RME case study schools: 5 schools, 2 RME teachers per school, 1 SLT/HoD, pupil focus group 4-8 pupils*
5. *All RME intervention school Y7 pupils in 2018/19: 7500 pupils*

Control

1. *All control schools: 59 schools*
2. *All control teachers: 118 teachers*
3. *Control teacher interview sample: 7 teachers*
4. *All control school Y7 pupils in 2018/19: 7500 pupils*

Sampling

To maximise the number of schools from which additional data will be collected, the aim will be for the interview, case study and where applicable the RME design schools samples to be mutually exclusive. However, this aim will be secondary to achieving intended data collection.

Samples will be informed by stratification based on hub,and using random selection.

Selection of which teachers to interview in the schools samples will be at random, choosing from teachers originally nominated to participate in RME or a nominated substitute if necessary.

If a school declines to participate, a reserve school will be identified that preserves the matrix criteria. If a teacher in an interview school declines or is unable to participate the other teacher will be invited to participate. If both teachers decline or are unable to participate then an alternative school will be identified.

If new teachers join the project in a school in which previously a teacher was interviewed, the new teacher will be invited to participate.

---

[15] Note that this assumes that the Y7 teachers go on to teach Y8 classes in all cases. MMU are anticipating the need for catch up CPD for teachers joining the project, thus the actual number of teachers surveyed is likely to be greater than 120. In addition, in the MoU it states that other teachers in the departments will be invited but not required to take the survey.

Overview of methods, rationale and analysis

School contextualising data

Data collected prior to randomisation from schools and from school census data will be used for contextualising and interpreting findings.

Teacher surveys (online)

Teacher surveys will collect data on teacher practices and beliefs.

Questions will be devised with consideration of the features of RME pedagogy, the intervention theory of change: that is related to multiple strategies, mathematising, redefining progress, representations, classroom talk, and use of context. Additionally, more generic beliefs about mathematics teaching or preference for practices will be surveyed, as well as efficacy in relation to RME related practices. Where appropriate, item design will draw on existing questions, for example the OECD TALIS survey (OECD 2014) and other appropriate instruments (see below, p.35, on the design process).

Teacher surveys will be administered electronically using Qualtrics software and use will be fully GDPR compliant.

Survey 1 in September 2018, will focus on teacher beliefs and practices in teaching mathematics and information on any prior use of *Making Sense of Maths* textbooks or of other RME materials.

Survey 2 (summer 2019), and survey 3: (summer 2020) will be tailored to the two trial conditions in the following ways:

- In RME schools the teachers originally nominated (or replacements) will complete a survey (30 minutes) on their mathematics teaching including in RME lessons, issues that might have affected implementation, attitudes to the value of project activities and materials in supporting effective teaching and professional development, and triangulation of data on implementation. Other teachers in the department will be invited (but not required) to complete the survey.
- In control schools the two teachers originally nominated to take part in the project will complete a short survey (20-minutes) on their mathematics teaching and issues that might affect pupil performance in mathematics in their school. This will assess the control condition; monitor programme differentiation and check for spill over.

Survey data will be downloaded and matched using SPSS and Excel software. Descriptive statistics will be generated from survey data to compare RME intervention teachers and control sample teachers and to identify change in RME intervention teacher responses over time. Relationship to fidelity data and variation in implementation will be analysed.

Collection of implementation data

Implementation data will be collected in two ways:

Directly from RME teachers, on the use of materials and engagement in other intervention activities. Information will be collected through short surveys (5-10 minutes to complete) completed via participants' handheld devices with a paper back up option at or after CPD days. These short surveys will record, for example,

1. how far they progressed through a module;

2. to what extent did they use the approaches in the teacher guide;

and 3. did they do the gap task.

1. , Surveys will be timed to coincide with CPD days as these coincide with when participants will have completed use of modules if they are compliant with the RME protocols. In order to maximise completion, teachers absent from CPD days will be emailed to encourage responses.
2. Collection of RME CPD attendance data from MMU.

Teacher interviews

These will be undertaken by telephone in spring 2019 and spring 2020. In addition to generic questions, interviewers will have access to interviewees' responses to teacher survey 1, and in the second interview, access to the transcript of the first interview and to teacher survey 2, where available. This will allow for respondent checking and enquiry about changes in practice.

- Intervention interviews

These will last 30-40 minutes and cover RME implementation, teacher views/beliefs about mathematics teaching and RME, and pupil responses. SHU will identify a group of 12 intervention schools and will invite one of the participating teachers from each of these schools to undertake a telephone interview of 30-45 minutes.

- Control school interviews

These will last 20 minutes and will focus on business as usual and also seek to identify any spill over issues and monitor programme differentiation.

Fieldworkers will complete summary notes recorded and managed in Excel in relation to research questions and sub-questions specific to the interview. Interviews will be professionally transcribed by a transcriber subject to confidentiality agreements. Transcripts will be managed and analysed using NVivo 11 software.

Case study visits

In depth longitudinal case studies will be developed from visits to 5 intervention schools. The purposes of the case studies are to:

- Understand in more depth how materials and pedagogical principles are implemented
- Seek explanatory insights into variability in implementation, including contextual factors such as senior leadership support
- Assess the extent to which RME practices are used in other lessons by RME teachers and whether RME teachers report practices or materials are being used by other teachers in intervention schools
- Triangulate and interpret pupil attitudes to mathematics data obtained from survey and pupil reports of their experience in mathematics
- Understand in more depth teacher change
- Identify issues relevant to scalability, such as support of senior leaders, contextual issues such as departmental and school characteristics
- Triangulate data for a sub sample in relation to fidelity and implementation obtained from self-reports (thus enhancing study reliability).

For the RME intervention schools, one visit will take place in autumn 2019 and one in summer 2020. During these visits, evaluators will:

- Observe at least one lesson taught by each participating teacher (on both visits) - on the first visit the observation will be of a lesson using RME material and on the second a lesson in which RME materials are not being used
- Interview both participating teachers (on both visits)
- Where possible, interview the head of department (or equivalent) and a member of the Senior Leadership team (on one or other visit)
- Conduct pupil focus groups with pupils experiencing the RME lessons (on one or other visit).

Interviews will be transcribed by professional transcribers bound by appropriate confidentiality agreements. Using field notes and interview records, fieldworkers will generate case study summaries including in relation to relevant research questions. In addition, transcripts and case study notes will be processed in NVivo 11 software. In addition, to considering each case separately, cross case analysis will be undertaken (Stake, 2013).

Pupil survey

In both intervention and control schools a pupil survey will be administered to all pupils on attitudes to mathematics and experience of mathematics teaching in summer 2020. The survey will collect data on pupil experience of practices and attitudes. This will happen shortly after the GL PTM is undertaken. Data will be collected on when the survey is administered. Both an online and paper option will be available.

The survey will consist of 15-25 items the majority or all of which will be in Likert form. Whenever appropriate, individual items will be selected or adapted from existing, relevant surveys of pupil attitudes and experience of mathematics teaching.

Descriptive statistics will be used to identify any significant differences between the intervention and control school samples and, within intervention schools, between pupils who have experienced RME teaching and those who have not. This will help to establish whether or not pupils experience a difference in teaching in RME compared to the control schools. In addition, items surveying attitudes towards mathematics will support interrogation of the theory of change.

Field visits to CPD days

CPD days will be visited on 4 occasions. The purpose of the visits will be to consider implementation of the professional development component and to understand more fully the intervention and teacher responses. Fieldworkers will observe and take notes on CPD activity, and make notes on teacher reflections on implementation including engagement in gap tasks.

Each visit will be summarised on a visit record and initial analytical notes made in relation to observation and relevant RQs, as well other issues of note.

MMU data summaries and project materials

MMU will provide the following data:

- Summary data on recruitment approaches
- Tri-annual updates to include: reflections on CPD events, summary of on-line activity; informal feedback, updates on any attrition and reasons for withdrawal, updates from design schools in relation to design research themes
- Project materials including module materials, CPD materials, gap task guidance

Data from these summaries will be analysed to principally to address RQ11 but will also inform design of instruments and IPE data collection, and to refine the project description in the final report.

Process for design of instruments

As noted above, the description of the intervention and programme theory of change was informed by an IDEA workshop (Humphrey et al. 2016). In addition, specific instruments will be designed with input from the MMU RME team. This includes teacher surveys, interview schedules, a pupil survey and observation rubrics.

The following process for instrument design will be followed. The SHU evaluation team will inform design of instruments from issues identified by MMU through updates on implementation. Drafts of instruments will be circulated for a round of substantive comments, consideration will be given to comments received and, where appropriate instruments will be revised.

Cost evaluation

Data on costs will be collected from the MMU team on the cost of CPD and materials. MMU will be asked to disaggregate any costs for development of materials or internal evaluation.

Because of variations in school costs for teacher release, the existence of such costs will be noted but not included in the cost per pupil calculation. Teachers will be asked to estimate any additional costs if additional to previous practice, for example additional or lower printing costs.

The intervention is for two years but costs over a three year period will be calculated to allow for comparison with other trials.

Table 8 Summary of IPE data collection methods and purposes

The table below summarises the IPE data collection, samples, purpose and details, and RQs addressed.

Method	Samples	When	Detail	Purpose
(i) Recruitment school information	Pre randomisation - intervention and control schools	Summer 2018	School information, e.g. MAT status, prior involvement in relevant CPD	Contextualising data
(ii) School census data	Pre randomisation - intervention and control schools	Summer 2018	Collected for the impact evaluation	Contextualising data
(iii) Teacher survey 1	Pre randomisation - nominated intervention and control school teachers	September 2018	Prior related use of practices and relevant CPD; beliefs and practices	Baseline on teacher beliefs and practices, see how these change over time
(iv) Teacher survey 2 and 3	End of year 7 and year 8 surveys	Summer 2019, 2020	Practices, relevant CPD, beliefs - for intervention triangulating data on implementation, for control assess control condition	Compare intervention and control schools, record longitudinal change in intervention schools
(v) RME teacher implementation data	RME teachers	At or near to 7 CPD days 2018-2020	Short survey on RME material use and RME CPD activity MMU attendance data	Assess implementation

(vi) RME Teacher interviews	RME teachers	Spring 2019 and Spring 2020	Interviews on RME implementation, change in practice, pupil engagement, RME culture	Assess implementation and theory of change
(vii) Control teacher interviews	Control teachers	Spring 2019 and Spring 2020	Interviews on business as usual, change in practice	Assess control condition
(viii) RME case studies	RME case study schools	Autumn 2019, Summer 2020	Interviews with teachers, observation, interviews with SMT, pupils	In depth understanding of implementation. Identify issues that might be relevant to scalability
(ix) Pupil survey	Survey of Y8 pupils	Summer 2020	Pupil experience of lessons and attitudes to mathematics	Triangulates data on teacher account of practices; assess RME culture
(x) CPD observations	Visits to CPD days	See timeline below for dates	Participant observation of CPD activities	Assess implementation of PD component
(xi) MMU data/reflection summaries and analysis of project materials	Collected direct from the MMU team	March 2019, July 2019, Dec 2019, March 2020, July 2020	Summary of data/analysis from MMU from delivering CPD and work with design schools. Analysis of project materials	Triangulates, provides RME team perspective, supports design of instruments. Supports description of intervention in final report

Ethics and registration

The RME evaluation was submitted to the SHU ethics committee and received approval in July 2018 (SHU Ethic Review ID: ER6803657).

In line with General Data Protection Regulation (GDPR) guidelines, because this evaluation research is being conducted on the legal basis of a public task, schools are not obliged to seek parental consent unless they prefer to do so for their own purposes. However, parents/carers will be informed that they can ask for their child’s data to be excluded at any stage of the trial. We suggested relevant phrasing for the parent/carer communication sent out by schools in September 2018. This relates to participation in the trial, linking to NPD data and participation in the GL PTM test in June 2020. Separate permission will be sought from parents/carers for the IPE data collections detailed in this document.

Personnel

MMU RME team

Principal Investigator: Yvette Solomon, Professor of Education

RME CPD lead: Sue Hough, Senior Lecturer in Mathematics Education

RME Design lead: Steve Gough, Senior Lecturer in Mathematics Education

Project manager: Vinay Kathotia, Senior Research Associate

RME trainers: Jo Kennedy, Senior Lecturer in Mathematics Education; Fiona Haniak-Cockerham, Senior Lecturer in Mathematics Education; Franke Eade, former Principal Lecturer in Mathematics Education; Marisa Bartoli, Laurus Trust Director of Mathematics

Project administrator: Jo Dennis until November 2018.

Research assistant and administrative support: Kate O'Brien

SHU Evaluation team

Principal Investigator and impact evaluation lead: Sean Demack, Principal Research Fellow

Co-Investigator and IPE lead: Professor Mark Boylan, Professor of Education

Project manager: Claire Wolstenholme, Research Fellow

Trial statistician and data manager: Dr Martin Culliney, Research Fellow

IPE case study lead and field worker: Dr Emma Rempe-Gillen, Senior Lecturer Mathematics Education

IPE fieldworker: Dr Gill Adams, Senior Lecturer Mathematics education

Project administrator/data entry: Judith Higginson, Administrator

Risks

Table 9 Assessment of risks to the intervention and evaluation

Risk	Assessment	Avoiding risk and controlling for its effect
Intervention insufficient or time too short to impact on pupil attainment.	Likelihood: Medium Impact: High	The trial spans two academic years to allow time for impact and is designed with relatively high statistical sensitivity (ITT MDES estimate = 0.18-0.24 depending on the degree of class-level clustering of the GL PTM).
Failure to recruit target of 120 non-selective	Likelihood: Medium Impact:	Reviewed regularly through the recruitment period and if recruitment is falling short of the 120 'ideal' school target recruitment will be expanded to include other types of schools (i.e. middle schools). The impact of this depends upon the

established secondary schools	Medium to High	number of 'ideal' schools in the trial - the smaller this is, the greater the impact on the sensitivity of the trial.
School drop-out / Attrition: Not participating in GL PTM test	Likelihood: Medium Impact: High	We propose that **all pupils** in the relevant year groups take PTM and all schools will receive a report on PTM attainment for their Y8 cohort in 2020 and this will act as an incentive for schools to participate in the GL PTM. Additionally, following the conclusion of the trial, control schools that participated in the data collection activities will be given a payment of £1000. MMU/EEF are funding an information **package for control schools**. This will be a one-day event that members of staff from each **control school** would be able to attend post testing. Depending on demand, this may happen in more than one location. Control schools will have full access to materials and guidance at the end of the trial. Control schools will be encouraged to remain in the trial and complete the post test. If a school indicates they intend to drop out, the head teacher will be contacted directly to ask to confirm withdrawal, given that they are a signatory to the MoU.
School not participating in IPE data collection	Likelihood: Medium Impact: Medium	Establishing a **positive/ongoing relationship with schools** throughout to ensure a quality flow of data/information that accurately informs both the process evaluation and post-test data. We will • provide information for delivery by MMU at the delivery inception CPD days to: introduce the evaluation, outline the key milestones and intended data-collection points (for both the testing and process evaluation) • ensure that all requests for data/time/information are proportionate and that burden on schools is minimised e.g. by providing standardised templates wherever possible, ensuring a high level of flexibility to accommodate school preferences for interviews/school visits etc. IPE response rates in two recent comparably sized trials MRP (DfE) and Dialogic Teaching (EEF) across the trials 71 and 74% intervention pupil post-tests and 78 and 84% control respectively.
Issues around co-ordinating and maintaining relationships with multiple partners	Likelihood: Low Impact: High	SHU has previously undertaken a number of previous evaluations involving MMU delivery (with both positive and no impact effects found). In addition the team has worked successfully with GL Assessment and Randstad Invigilators on a number of recent projects.
Slippage and deadlines not met	Likelihood: Low Impact: High	Effective project management, monitoring, risk management and quality assurance processes. Appropriate communication strategy determined early between SHU and EEF project managers to anticipate and head-off emerging problems.
Data Protection and Ethical Issues	Likelihood: Low Impact: High	Robust data protection and ethical procedures are in place at SHU and MMU and data sharing protocols will be established.
Staff absence (e.g. due to	Likelihood: Low	The evaluation team has experienced statisticians and project managers that can cover in the event of staff absence or

long term illness)	Impact: High	departure.

Timeline

The table below provides the timeline for the intervention.

Table 10 Schedule of intervention and evaluation activities

Date	Activity	Who?
January 2018	Recruitment opens	MMU
June - July 2018	Collection of initial school data	MMU
28 Sept 2018	Recruitment deadline - all schools to return signed MoU, provide school and pupil data, complete baseline teacher survey	Schools/MMU/SHU
5 Oct 2018	Randomisation of schools to intervention/control group	SHU
Oct 2018	RME CPD Day 1	MMU/teachers
Nov 2018	Design of research instruments - CPD day surveys, teacher interview schedules	SHU
Nov-Dec 2018	RME teaching - module 1 Number	RME teachers
Dec 2018	Identification of sub sample for additional evaluation activities RME CPD Day 2 CPD day survey 1 (focus on Number module)	SHU/MMU (design schools) MMU/teachers SHU/teachers
Jan 2019	Statistical analysis plan drafted	SHU
Jan-Feb 2019	RME teaching - module 2 Geometry	RME teachers
Feb 2019	RME CPD Day 3 CPD day survey 2 (focus on Geometry module)	MMU/teachers SHU/teachers
March-April 2019	RME teaching - module 3 Proportional Reasoning (PR) MMU visits to design schools Teacher interviews intervention and control	RME teachers MMU/schools SHU/teachers
April 2019	RME CPD Day 4 (introduces two modules) CPD day survey 3 (focus on PR module) Design/refine end of year teacher survey	MMU/teachers SHU/teachers SHU
May-June 2019	RME teaching - module 4 (Data or Algebra) MMU visits to design schools	RME teachers MMU/schools
June-July 2019	RME teaching module 5 (Algebra or Data) RME CPD day 5 (introduces two Y8 modules) CPD day survey 4 (focus on modules 4,5) Teacher survey - intervention and control schools	RME teachers MMU/teachers SHU/teachers SHU/teachers
Sept 2019	Design of schedules for evaluation case study visits 1 day CPD catch up for Y8 teachers joining the project RME teaching (Y8) - module 6 Number	SHU MMU/new teachers RME teachers
Oct 2019	RME teaching (Y8) module 7 PR RME CPD day 6 (introduces two modules) CPD day survey 5 (focus on modules 6,7)	RME teachers MMU/teachers SHU/teachers
Nov-Dec 2019	RME teaching (Y8) module 8 (Geometry or Data) SHU visit 1 to case study schools MMU visits to design schools Update teacher-class-pupil data (update 1)	RME teachers SHU/schools MMU/schools SHU/schools
Dec 2019-Jan 2020	RME teaching (Y8) module 9 (Data or Geometry) MMU visits to design schools	SHU/teachers MMU/schools
Feb 2020	RME CPD day 7 CPD day survey 6 (focus on modules 8,9)	MMU/teachers SHU/teachers
March-April 2020	RME teaching (Y8) module 10 Algebra SHU telephone interviews - intervention and control	RME teachers SHU/teachers

	MMU visits to design schools	MMU/schools
April-May 2020	RME CPD day 8 CPD day survey 7 (focus on module 10 Algebra) Update teacher-class-pupil data (update 2) MMU visits to design schools	MMU/teachers SHU/teachers SHU/schools MMU/schools
June-July 2020	GL Progress Test in Maths all Y8 pupils Y8 pupil survey Conduct post-intervention survey for all teachers Second visit to intervention case study schools	SHU/schools SHU/schools SHU/teachers SHU/schools
July 2020	Schools receive GL test outcomes Control schools receive RME materials and payment, and invited to attend one day event on RME	SHU/GL MMU
June/July 2020	Outcome assessment administration and marking	SHU (with contractors)
Summer 2020	Analysis of lessons observed in case study visit 2 Analysis and report writing	SHU SHU
October 2020	Draft report submission	SHU
Spring 2021	Report publication	SHU

REFERENCES

Allen, R., Jerrim, J., Parameshwaran M. & Thompson, D. (2018) Properties of commercial tests in the EEF database. EEF Research Paper 1, 2018

Barmby, P., Dickinson, P., Hough, S., and Searle, J. (2011) Evaluating the impact of a Realistic Mathematics Education project in secondary schools. *Proceedings of the British Society for Research into Learning Mathematics* 31(3) 47-52.

Boaler, J. (2002). Experiencing school mathematics: Traditional and reform approaches to teaching. New Jersey: Lawrence Erlbaum Associates.

Boaler, J. (2002). Paying the price for" sugar and spice": Shifting the analytical lens in equity research. *Mathematical Thinking and Learning,* 4(2-3), 127-144.

Boaler, J. (2003) Studying and capturing the complexity of practice – the case of the 'Dance of Agency'. In Pateman, N., Dougherty, B. and Zilliox, J. (Eds. 2003) *Proceedings of the 27th annual conference of the International Group for the Psychology of Mathematics education.* (Vol. 1, pp. 3-16) Honolulu, HI: PME.

Boaler, J., and Wiliam, D. (2001). "We've still got to learn!" Students' perspectives on ability grouping and mathematics achievement. In P. Gates (Ed.), *Issues in mathematics teaching*. London: Routledge Falmer.

Boaler, J., Wiliam, D., and Brown, M. (2000). Students' experiences of ability grouping-disaffection, polarisation and the construction of failure. *British Educational Research Journal, 26*(5), 631-648.

Boylan, M. and Demack, S. (2018) Innovation, evaluation design and typologies of professional learning. *Education Research Journal.* Vol. 60 (3) pp336-356.

Boylan, M., Demack, S., Willis, B., Stevens, A., Adams, G and Verrier, D. (2015a) *Multiplicative reasoning professional development programme: Report.* London: DfE.

Boylan, M., Demack, S., Willis, B., Stevens, A., Adams, G and Verrier, D. (2015b). *Multiplicative reasoning professional development programme: Technical Report.* London:

DfE.

Boylan, M., Demack, S., Wolstenholme, C., Reidy, J. and Reaney-Wood, S. (forthcoming) ScratchMaths Evaluation Report. The 2017 protocol for this evaluation is here: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Round_6-_Scratch_maths_amended.pdf

Boylan, M. and Jay, T. (2017) Independent evaluation of 'Investigating the impact of Realistic Mathematics Education approach on achievement and attitudes in Post-16 GCSE mathematics resit classes' In 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*. Nuffield.

De Corte, E., Op 't Eynde, P., and Verschaffel, L. (2002). "Knowing what to believe": The relevance of students' mathematical beliefs for mathematics education. In B. Hofer and P.

De Lange, J. (1996). Using and Applying Mathematics in Education. In A.J. Bishop, et al. (eds) *International handbook of mathematics education, Part One*. 49-97. Kluwer Academic 101

Dickinson, P., & Eade, F. (2005). Trialling realistic mathematics education (RME) in English secondary schools. *Proceedings of the British Society for Research into Learning Mathematics,* 25(3), 1-13.

Dickinson, P., and Hough, S. (2012). *Using realistic mathematics education in UK classrooms*. Centre for Mathematics Education, Manchester Metropolitan University, Manchester, UK.

Dickinson, P., Eade, F., Gough, S. and Hough, S. (2010) Using Realistic Mathematics Education with low to middle attaining pupils in secondary schools. In Joubert, M. (Ed.) *Proceedings of the British Society for Research into Learning Mathematics* April 2010

Drury, H. (2014). *Mastering mathematics. Oxford:* Oxford University Press.

EEF. (2017) *Improving mathematics in Key Stages Two and Three.* London: EEF.

Fosnot, C.T. and Dolk, M. (2002). *Young Mathematicians at Work: Constructing Fractions, Decimals, and Percents*. Portsmouth, NH: Heinemann.

Francis, B., Archer, L. Hodgen, J., Pepper, D., Sienna, B. and Travers, M-C (2017) Exploring the relative lack of impact of research on 'ability grouping' in England: a discourse analytic account, *Cambridge Journal of Education*, 47:1, 1-17

Gravemeijer, K., Van den Heuvel, M. and Streefland, L. (1990). *Contexts, Free Productions, Tests, and Geometry in Realistic Mathematics Education*. Utrecht: OW and OC.

Hedges, L.V. & Rhoads, C. (2010) Statistical Power Analysis in Education Research. NCSER 2010-3006. Available at https://ies.ed.gov/ncser/pubs/20103006/pdf/20103006.pdf

Hodgen, J., Foster, C., Marks, R., and Brown, M. (2018). *Improving mathematics in key stages two and three: evidence review.*

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*. MMU/Nuffield.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016) Implementation and process evaluation (IPE) for interventions in education settings: A synthesis of the literature. (Manchester, Manchester Institute of Education). https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Review_Final.pdf.

Jerrim, J., and Choi, Á. (2014). The mathematics skills of school children: how does England compare to the high-performing East Asian jurisdictions?. Journal of Education Policy, 29(3), 349-376.

Kelcey, B., Spybrook, J., Phelps, G., Jones, N. & Zhang, J. (2017) Designing large scale multisite and cluster randomized studies of professional development. The Journal of Experimental Education, 85(3) pp389-410

Lewis, G. (2011) Mixed methods in studying the voice of disaffection with school mathematics in Smith, C.(Ed.) *Proceedings of the British Society for Research into Learning Mathematics 31(3)* November 2011

Mason, J. (2002). *Researching your own practice: The discipline of noticing*. Psychology Press.

Meelissen, M., and Luyten, H. (2008). The Dutch gender gap in mathematics: Small for achievement, substantial for beliefs and attitudes. *Studies in Educational Evaluation,* 34(2), 82-93.

Nardi, E. and Steward, S. (2003) Is mathematics T.I.R.E.D.? A profile of quiet disaffection in the secondary mathematics classroom. *British Educational research Journal 29 (3)*: 345-67

National Centre for Excellence in the Teaching of Mathematics (NCETM) (2014). Mastery approaches to mathematics and the new national curriculum. https://www.ncetm.org.uk/public/files/19990433/Developing_mastery_in_mathematics_october_2014.pdf

Organisation for Economic Co-operation and Development. (2014). *TALIS 2013 results: An international perspective on teaching and learning.* Paris: OECD.

PISA (2000, 2006, 2009). *Programme for International Student Assessment* (OECD), retrieved April 2017 from http://www.oecd.org/pisa/

PISA (2015 *Programme for International Student Assessment* (OECD), retrieved January 2019 from http://www.oecd.org/pisa/

Romberg, T. and Shafer, M. (2005). T*he longitudinal/cross-sectional study of the impact of teaching mathematics using Mathematics in context on student achievement: Implications and con*clusions. Wisconsin: Wisconsin Center for Educational Research. http://micimpact.wceruw.org/

Searle, J. and Barmby, P (2012) *Evaluation Report on the Realistic Mathematics Education Pilot Project at Manchester Metropolitan University.* Durham: Durham University.

Spybrook, J., Shi, R., Kelcey, B. (2016) Progress in the past decade: an examination of the precision of cluster randomized trials funded by the U.S. Institute of Education Sciences. International Journal of Research and Method in Education. 39 (3) pp255-267

Stake, R. E. (2013) *Multiple case study analy*sis (New York, Guilford Press).

Streefland, L. (1985). Wiskunde als activiteit en de realiteit als bron [Mathematics as an activity and reality as a source]. *Nieuwe Wiskrant, 5*(1), 60-67.

Swan, M. (2006) Collaborative learning in mathematics: a challenge to our beliefs and practices Leicester: National Institute of Adult Continuing Education.

TIMSS (1999, 2007, 2010). Trends in International Mathematics and Science Study. TIMSS International Study Centre, retrieved April 2017 from https://timssandpirls.bc.edu/

Treffers, A. (1987) *Three dimensions: A model of goal and theory description in mathematics instruction: The Wiskobas project*. Dordrecht: Reidel.

Van den Heuvel-Panhuizen, M. (2003) The didactical use of models in realistic mathematics education: an example from a longitudinal trajectory on percentage. *Educational Studies in Mathematics,* 54 (1), 9-35.

Van den Heuvel-Panhuizen, M., & Drijvers, P. (2014). Realistic Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 521-525). Dordrecht, Heidelberg, New York, London: Springer.

Webb, D., Boswinkel, N. and Dekker, T. (2008) Beneath the tip of the iceberg: Using representations to support student understanding. *Mathematics teaching in the middle school,* 14 (2), 110-113.

Weiss, C. H. (1997). How can theory-based evaluation make greater headway?. *Evaluation review,* 21(4), 501-524.

Yackel, E. & Cobb,P. (1996) Sociomathematical Norms, Argumentation, and Autonomy in Mathematics. *Journal for Research in Mathematics Education*, 27 (4). 458-477.

Appendix I Realistic Maths Power Analysis

From Kelcey et al (2017), the Minimum Detectable Effect Size (MDES) for a 3-level CRT is

$$MDES_{3LCRT} \sim M_{K-L-2}\sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{ICC_{sch}(1-R^2_{sch})}{K}+\frac{ICC_{class}(1-R^2_{class})}{JK}+\frac{(1-ICC_{sch}-ICC_{class})(1-R^2_{pup})}{nJK}}$$

From Spybrook et al., (2016), the MDES equation for a 4-level MSCRT but assuming zero effect size variability across clusters but also including covariate explanatory power at class and pupil levels is:

$$MDES_{4LMSCRT} \sim M_{(M(K-L-2)}\sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{ICC_{sch}(1-R^2_{sch})}{MK}+\frac{ICC_{class}(1-R^2_{class})}{MKJ}+\frac{(1-ICC_{sch}-ICC_{class})(1-R^2_{pup})}{MKJn}}$$

It can be useful to re-organise this equation following Hedges & Rhoads (2010)…

$$MDES_{4LMSCRT} \sim M_{(M(K-L-2)}\sqrt{\frac{1}{P(1-P)MKJn}}\sqrt{1+(Jn-1)ICC_{sch}+(n-1)ICC_{class}-\left[R^2_{pup}+\left(JnR^2_{sch}-R^2_{pup}\right)ICC_{\ Sch}+\left(nR^2_{Class}-R^2_{pup}\right)ICC_{\ Class}\right]}$$

Where…

- P is the proportion of schools who receive the intervention (=0.50)
- R^2_{Sch} is the school-level covariate explanatory power (=0.49)
- R^2_{class} is the class-level covariate explanatory power (=0.49)
- R^2_{pup} is the pupil-level covariate explanatory power (=0.49)
- ICC_{Sch} is the school level Intra Cluster Correlation coefficient (=0.15 to 0.20)
- ICC_{class} is the class level Intra Cluster Correlation coefficient (=0.05 to 0.50)
- M is the number of geographical sites (=6)
- K is the number of schools per site (=20)
- J is the number of classes per school (=3)
- n is the number of pupils per class (=25)
- L is the number of school level covariates (=9)
- $M_{(M(K-L-2)}$ is the t-distribution multiplier with M(K-L-2) (54) degrees of freedom. Assuming a two-tailed test with a statistical significance of 0.05 (α/2=0.025) and statistical power of (1-β=0.80). M_{54} = 2.8532.

Therefore, a range of MDES estimates for different strengths of clustering at the school and class level can be calculated; below are two examples based on extremes of clustering at school/class levels.

Weakest clustering ($ICC_{Sch} = 0.15;\ ICC_{class} = 0.05$); $MDES = 2.8532\sqrt{0.00302} = 0.157 \sim 0.16$

Strongest clustering ($ICC_{Sch} = 0.20;\ ICC_{class} = 0.50$); $MDES = 2.8532\sqrt{0.0063} = 0.227 \sim 0.23$

Appendix II Additional IPE tables

Table 1 Indicative content of CPD days

Training days	Activities
Day 1	Introduction to the project. Introduction to RME and to the Y7 Number module, including modelling an RME Number lesson and planning for teaching. Planning for gap task 1: Observing and diagnosing students' mathematical issues (pre-test and video of students' working on number problems).
Day 2	Reflection on teaching of RME Number module. Analysis of gap task 1: Reviewing videos and scripts of students working on problems. Introduction to Y7 Geometry module and planning for teaching. Discussion of RME principles and approaches. Planning for gap task 2: Noticing and working on specific pedagogic strategies (teachers video each other).
Day 3	Reflection on teaching of RME Geometry module and on gap task 2. Introduction to Y7 Proportional Reasoning module and planning for teaching. Further development of RME principles and approaches. Planning for gap task 3: annotation of teaching guides.
Day 4	Reflection on teaching of RME Proportional Reasoning module and on gap task 3. Introduction to two Y7 modules, Algebra and Data, and planning for teaching. Further development of RME principles and approaches. Planning for gap task 4.
Day 5	Reflection on teaching of RME Algebra and Data modules and on gap task 4. Introduction to two Y8 modules (for teaching in Autumn 2019), Number and Proportional Reasoning, and planning for teaching. Further development of RME principles and approaches. Planning for gap task 5.
Catch up CPD Day for new teachers	The day is for teachers new to Year 2 of the project. It will cover 1) an introduction to RME. 2) An overview of the mathematical content of all 5 Y7 modules, initial (two) Y8 modules, and related RME principles and approaches. 3) Pedagogic practices, with a focus on the strategies that have been covered via the gap tasks thus far. 4) Walk-through of RME web-based support.
Day 6 (Year 8)	Reflection on teaching of Y8 RME modules, Number and Proportional Reasoning, and on gap task 5. Introduction to two Y8 modules, Data and Geometry, and planning for teaching. Further development of RME principles and approaches. Planning for gap task 6.
Day 7 (Year 8)	Reflection on teaching of Y8 RME modules on Data and Geometry, and on gap task 6. Introduction to the final Y8 module, Algebra, and planning for teaching. Further development and refinement of RME principles and approaches. Planning for gap task 7: adapting existing curricular and teaching materials for an RME approach.
Day 8 (Year 8)	Reflection on teaching of Y8 RME Algebra module and discussion and reflection on gap task 7. Discussion of implications of RME approaches for long-term learning and future practice as individual teachers and as a department.

Table 2 RME intervention curriculum materials

Module (Roughly 2 weeks of lessons)	Two week module to be taught within month(s) indicated below	Content focus
Y7 Number	Nov 2018	Diagrammatic representations of fractions Comparing fractions – various strategies Adding and subtracting fractions Finding fraction of an amount
Y7 Geometry	Jan or Feb 2019	Finding area of various shapes including rectangles, triangles, parallelograms, trapezium and compound shapes. Looking at relationships between units of area
Y7 Proportional Reasoning	March or April 2019	Diagrammatic representations of fractions Finding percentage of an amount Estimating percentages Expressing one amount as a percentage of another Percentage change Proportional quantities
Y7 Data	May or June 2019	The dot plot Pictograms Mean, median, mode and range for discrete data Comparing and interpreting statistical data
Y7 Algebra	May or June 2019	Letters and symbols for unknowns +, –, x, ÷ in algebra Collecting like terms, expand single brackets, factorise Develop algebraic thinking for solving equations
Y8 Number	Sept or Oct 2019	Multiplication and division strategies Ratio notation. Dividing in a given ratio
Y8 Proportional Reasoning	Sept or Oct 2019	Comparing proportional quantities Speed, distance, time Solve problems involving proportion Working with fractions and decimals
Y8 Geometry	Nov, Dec 2019 or Jan 2020	Finding area and perimeters for circular shapes Finding volume and surface area of 3D solids Pythagoras Theorem
Y8 Data	Nov, Dec 2019 or Jan 2020	Drawing pie charts Interpreting a variety of data representations including stacked bar charts Strategies for finding the mean, median and mode for grouped data
Y8 Algebra	March or April 2020	Developing the concepts of gradient and intercept Working with formulae and graphs Re-arranging formulae

MMU Design Schools Research Questions:

1. What elements of RME pedagogic practices are evident (sharing student strategies, encouraging students to 'mathematise' and offer viewpoints, supporting problem-solving, working with context)? How does the teacher use the lesson time (time on context, whole class discussion, individual work etc.), and how do they use the materials (e.g. time spent on context versus formal slides, elaborations and departures, time spent on available lessons in a block)?

2. How do teachers relate their classroom practice to the PD input they have received? Do they identify any changes from before PD?

3. What elements of an RME classroom culture are evident in terms of student behaviour (engaging with context, approaching the board, raising issues/volunteering insights, discussing strategies, disagreeing, explaining thinking)?

4. What evidence is there of an RME approach in student work? Do students use RME methods e.g. the bar, ratio table, drawing diagrams? How do teachers and students use the bar?

5. How do teachers undertake the gap tasks and why?

APPENDIX III: Fidelity components (to be agreed/finalised in the SAP)

Table 3

Component	Data	Collection	Maximum	Fidelity level (min)	Fidelity level (high)	Criteria
Teacher level						
T1. Professional development attendance number of days per teacher	Primary: MMU attendance list Secondary: Triangulation/assessment of reliability through end of year teacher survey	MMU SHU online Intervention survey	8	4	7	Half day or more counted as 1
T2 PD attendance at modules that have number content (what are currently called -Y7 and Y8 number, Y7 and Y8 proportional reasoning)	MMU attendance lists Secondary: Triangulation/assessment of reliability through end of year teacher survey	MMU SHU online Intervention survey	4	2	3	Half day or more counted as 1
T3. In school PD tasks - individual	Number of gap tasks completed	PD day survey - on-line hand held device paper option	7	3	5	What counts as 'completion of each gap task to be specified in the SAP
T4 Material use - curriculum time	Primary: Percentage of school lessons used based on number of maths lesson per week collected from PD day survey - based on 60 possible weeks RME could be taught Secondary: annual teacher survey for triangulation	PD day survey - on-line hand held device with paper option Survey for triangulation	20 weeks out of 60 - based on 2 weeks per 10 modules	17% of total maths available teaching time (50 % of RME 20 weeks)	27% of maths available teaching time (80% approx of RME 20 weeks)	Teacher self-report that a lesson as a RME module lesson counts as 1 lesson

T4 Materials use - blocks	Taught in a 'block'	PD day survey	10	5/6	8	What counts as a block to determined and recorded in the SAP
T5 Material use: module coverage	Number of modules and what materials used.	CPD day survey	10 modules	7	8	The amount of material/amount of teaching in each module to count as coverage to specified in the SAP
T6 Materials use number modules (currently Y7 number, Y8 numbers, Y7 proportional reasoning, Y8 proportional reasoning	Number of the 4 modules focused on number	CPD day survey	4 number modules	3	4	The amount of material/amount of teaching in each module to count as coverage to specified in the SAP

School level						
S1 PD attendance - paired attendance	Primary: MMU attendance list Secondary: Triangulation/assessment of reliability through end of year teacher survey	MMU SHU online Intervention survey	8	3	6	Two teachers attend at least Half day each. Note if more than two teachers participating, additional teacher's attendance not considered.
S2 Collaboration on the gap tasks	Teachers working together on the gap tasks, co-planning etc.	CPD day surveys	7	3	5	What constitutes collaboration to be identified in the SAP.