Department for Education

Multiplicative reasoning professional development programme: evaluation

Technical report & detail of evaluation findings

Research report

June 2015

Mark Boylan, Sean Demack, Ben Willis, Anna Stevens, Gill Adams and Diarmuid Verrier – Centre for Education and Inclusion Research (CEIR), Sheffield Hallam University

Contents

List of figures

List of tables

1. Introduction

This technical report accompanies the Main Report of the Evaluation of the Multiplicative Reasoning Project. It provides further detail on the background, structure and organisation of the project, the evaluation design and details of methodology. It also has details of both the impact analysis and the process evaluation that underpins the findings reported in the main report. It also includes the school case studies.

2. Background

The Multiplicative Reasoning Project (MRP) was commissioned by the Department for Education (DfE) to:

- improve teaching of mathematics at Key Stage 3 (KS3) with only 38% of lessons rated as good or better by OFSTED and 12% inadequate[1];
- design professional development opportunities for less experienced and non-specialist teachers who are often allocated to KS3 classes[1]
- address the relative drop in rate of progress of English pupils compared to comparator nations with higher than average levels at end of primary school but lower than average at end of KS3[2]

The particular focus on multiplicative reasoning follows from its importance within mathematics. The multiplicative nature of the reasoning used to solve problems arising from proportional situations often involves an understanding and application in different ways of fractions as well as decimals, percentages, ratio and proportion (see Annex A and Annex C). Together these constitute a significant part of the mathematics curriculum and in addition proportionality is connected to many other curriculum areas and has applications across other subjects. Pupils (and adults) fail to move on from additive structures and this can lead to many misconceptions and errors in subsequent mathematical study. The rationale for a focus on multiplicative reasoning and the project were informed by (see Annex A):

- the recognition for at least 30 years that pupils often fail to understand fundamental concepts of multiplication and division, ratio, proportion and rates of change[3]
- the need to reverse the trend of declining in understanding of multiplicative reasoning compared to pupils 30 years ago[4].
- a view that where techniques and methods are learnt they are often misapplied[5]

---

[1] OFSTED (2012) *Made to measure: messages from inspection evidence*. OFSTED

[2] Mullis I. et al. (2012). *TIMSS 2011 International Results in Mathematics*. *International Association for the Evaluation of Educational Achievement*

[3] Hart, K.M. (Ed) (1981), *Children's Understanding of Mathematics*: 11-16, John Murray

[4] Brown, M. et al. (2010) Attitudes, gender and attainment: evidence from a large scale survey in England. Paper presented at the *34th Conference of the International Group for the Psychology of Mathematics Education*, Belo Horizonte, Brazil:

- an awareness that whilst attempts have been made to address these issues, through National Strategies and other initiatives, there has likely been little genuine progress[6].
- the belief that interventions with a coherent approach and using relevant models can increase pupils' understanding of multiplicative reasoning[7]

The MRP was informed by previous effective forms of professional development undertaken by the National Centre of Excellence in Mathematics (NCETM), such as Mathematics Knowledge Networks[8] and the Primary Mathematics Hosts Schools Project[9].

After NCETM were commissioned to develop a project, discussions took place about how the project would be evaluated. The choice of an randomised control trial (RCT) as the evaluation methodology was influenced by current policy aims to develop evidenced informed educational practice[10].

---

[5] Watson, A., Jones, K. and Pratt, D. (2013), *Key Ideas in Teaching Mathematics: Research Based Guidance for ages 9-19*, OUP

[6] Hodgen, J., Brown, M., Küchemann, D., & Coe, R. (2010). *Mathematical attainment of English secondary school students: a 30-year comparison.* Paper presented at the British Educational Research Association (BERA) Annual Conference, University of Warwick.

[7] **See Increasing Student Competence and Confidence In Algebra and Multiplicative Structures (ICCAMs) project http://iccams-maths.org/**

[8] Gousetti, A., Potter J. and Selwyn N., 2011 Assessing the impact and sustainability of networks stimulated and supported by the NCETM. London: Londonknowledgelab. Url https://www.ncetm.org.uk/files/7979131/NCETM+Network+evaluation+report+(final)+amended+300611.pdf Accessed October 2012.

[9] Boylan, M. & MacNamara, A. (2013). The evaluation of the NCETM primary mathematics host schools project. CEIR

[10] Goldacre, B. (2013). Building evidence into education. http://media.education.gov.uk/assets/files/pdf/b/ben%20goldacre%20paper.pdf

3. The Multiplicative reasoning project

3.1 The project design

The project was designed primarily as a pilot professional development project that paid attention to the effects of teacher learning on pupils. It incorporated a number of both established approaches to professional development and ones that were more novel or had not been utilised in the same project. The specific aims and objectives of the project were related to the development of pupil understanding of the nature of multiplicative relationships and the appropriate use of models and algorithms to solve related problems. However, it was also intended that the structure and implementation of the Professional Development programme should provide a working model for future developmental programmes related to other mathematical content. Additionally, it was hoped that pupils who engaged with project activities would develop a more positive opinion of mathematics. Significant design features are provided in the main report.

3.2 Project aims and objectives[11]

Aims

As a pilot, the project aimed to provide a foundation from which further work nationally and within departments could be developed, and to explore a research informed orientation to MR professional development drawing on three different approaches.

The core purpose of the project was to improve the outcomes and experience for KS3 pupils in the area of multiplicative reasoning leading to wider benefits for their mathematics education.

As well as improved outcomes for pupils, the development programme attempted to support further outcomes including, at project inception:

- professional development for the teachers participating in the programme
- leadership development for the teachers co-leading each local development team
- curriculum and resource development resulting in teaching plans and resources (three teaching units for each KS3 year)

[11] See Technical Report Annex A

- the development of practices and resources that support teachers in being evidence informed practitioners
- the establishment of Mathematics Education Strategic Hubs (MESH) led by teaching schools that can sustain the model of development in the future

Objectives

Specific intended outcomes for teachers were

- to develop an understanding of the general nature of the difficulties experienced by pupils when learning about multiplicative structures

- to identify and understand the specific problems experienced by the pupils in the teachers' own schools when learning about multiplicative structures

- to develop their own subject knowledge regarding ways in which pupil understanding can be enhanced through the use of specific models and algorithms

- to have trialled and evaluated both generic and specific resources which can be used longer term and incorporated into schemes and plans

- to develop department-wide policies and strategies regarding effective pedagogies to employ when creating and teaching lessons and activities related to multiplicative structures

Specific outcomes for pupils were intended to be:

- to develop a deeper understanding of the nature of multiplicative structures

- to learn to use models and algorithms appropriately to solve problems involving multiplicative structures

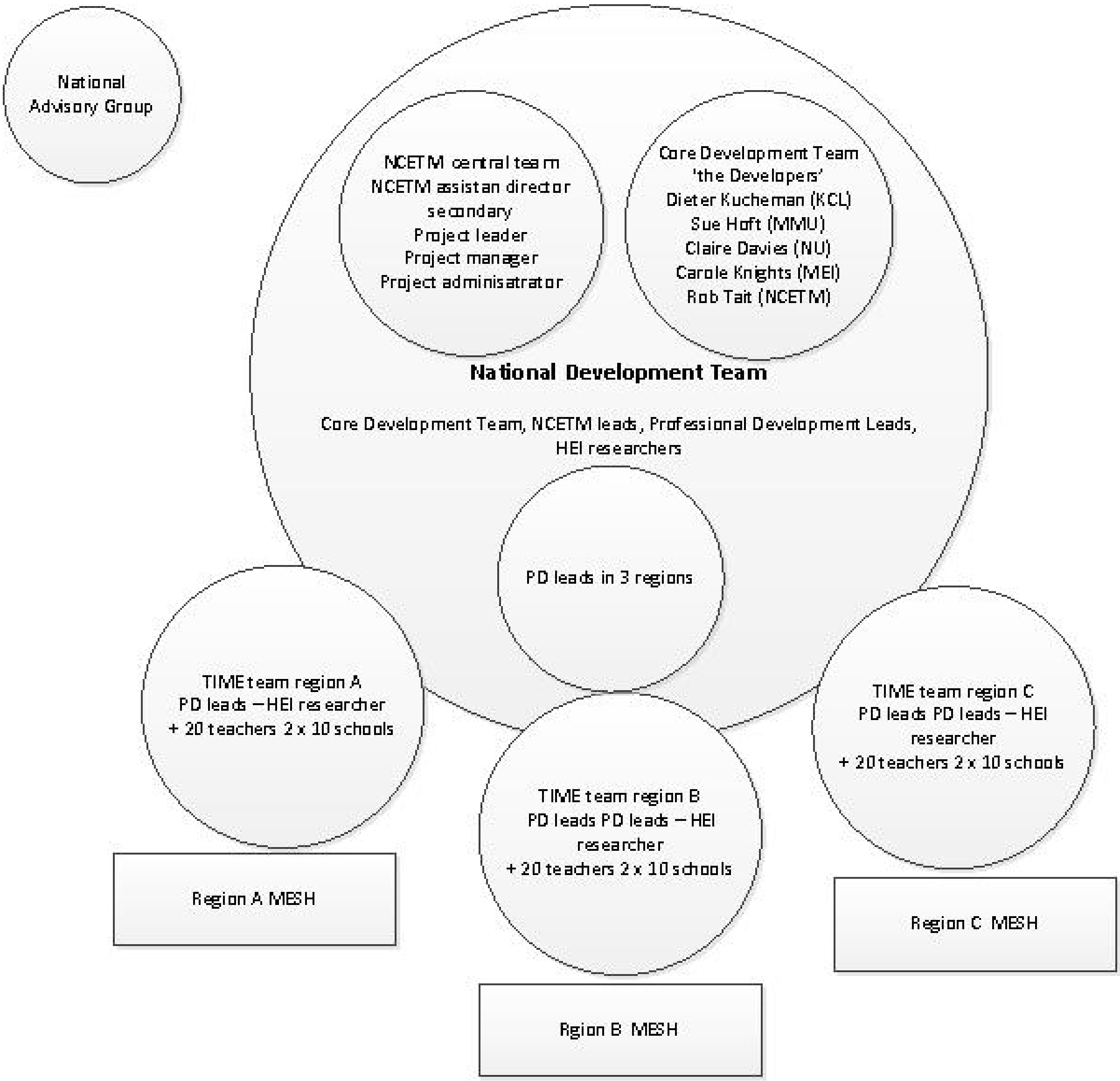
One of the project objectives was to develop department wide policies and strategies. In the original design for the evaluation, work by teachers within departments was emphasised and initially impact on pupil performance across departments was to be measured. In the event, the scope of the project was more clearly focused on outcomes for the teachers involved.

3.3 Project organisation and structure

3.3.1 Overview of organisation

The project had a national, regional and school level structure. The diagram below provides an overview of the organisation of the project as a whole.

Figure 1 Project organisation



3.3.2 Project roles and groups

In this section, we provide some additional detail about the intended remit and responsibilities of each of the key groups/organisations identified in the above diagram. Section 8.1 provides a narrative about the activity of these groups and roles in practice and a list of personnel is given in Appendix B.

National Advisory Group (NAG): The NCETM was supported by a NAG consisting of mathematics educators, officials and research experts, the NCETM secondary director and NCETM project lead.

NCETM Project Lead - To oversee the delivery of the project and act as a link between different roles/contributors and between the three TIME teams.

Core Development Team (CDT): The CDT consisted of three developers from three different universities, one from Mathematics in Education and Industry (MEI) and the NCETM project lead from NCETM. The developers were selected to bring different perspectives and approaches to mathematics pedagogy/curriculum development and were responsible for supporting the development of teaching plans, resources and the PD programme.

National Development Team (NDT): The NDT consisted of the CDT, the PD leads from each TIME team and university researchers. The NDT remit was to maintain an overview of the project and to plan in more detail and lead implementation of the professional development related to the resources created. The NDT finalised curriculum materials and professional development activities and planed the TIME team events.

University researchers - The project consisted of three Higher Education Institutes (HEI) researchers with a nominated HEI researcher allocated to a specific regional TIME team. The HEI researchers were responsible for providing a research informed perspective on MR, developing the teaching resources and contributing to lesson study. The HEI researchers were part of the National Development Team.

Professional development (PD) Lead: There were two PD leads per TIME team and a total of six across the project. PD Leads were expected to possess: comprehensive understanding of effective pedagogy within the KS3 mathematics curriculum, experience of leading professional development with mathematics teachers (ideally at KS3), substantial experience of teaching mathematics at KS3 and knowledge of recent and imminent developments within KS3 mathematics. PD leads had a variety of backgrounds, some being new to professional development leadership, others having worked in a variety of consultant/school improvement roles previously.

Following their recruitment PD leads undertook the following actions/responsibilities:

- *Participating in the five national development workshop days*: Tasks included working with the CDT to plan the structure and content of the TIME team workshop days; providing feedback and evaluation from these days as required and considering successes and challenges of the project and models for sustainability.
- *Leading six TIME team workshop days*: Tasks included working with the MESH co-ordinator on the practical arrangements for the workshop days; facilitating workshop days, collaborating with the university researcher in supporting teachers to develop their own lesson study/pupil study practices
- *On-going support for the TIME team* via the online community by animating discussion and seeking support from the CDT when required.
- *Communicating with the project lead and developers*, either through visits by the lead or online discussion, raising any issues emerging from the TIME team's work or from project schools communicating any emerging needs regarding resource requirement and use.

Teachers Improving Mathematics Education (TIME) teams: There were 3 TIME teams within the project, located in three different regions that crudely gave some degree of geographical representation. Each team was intended to have 20 'core' teachers (2 each from 10 schools - representing different levels of experience in the pair) to represent differing experience with one teacher nominated as the Core Teacher (CT) and be led by two professional development leads and one university researcher. TIME teams met together for six workshop days spread throughout the course of the project. The TIME teams were intended to be collaborative, with teachers acting as fully engaged participants and not passive recipients.

Mathematics Education Strategic Hubs (MESH): The NCETM sought to involve the MESH to help co-ordinate the work of the TIME teams as seemed prudent given the proposed remit MESH have for providing support for high quality PD. The NCETM worked with a nominated MESH co-ordinator in the following four aspects:

- *Provision of a venue and facilities for TIME team workshops*: Organisation of practical arrangements of venue provision (typically a training facility within the school) and undertaking some resource copying as required for the workshops.
- *Liaison with the TIME team leadership group:* In conjunction with the NCETM helped recruit the PD leads and university researcher. Ongoing dialogue with PD Leads to ensure they had all the practical support required to run the workshops and with the university researcher to strengthen links to support future work. In addition PD Leads were expected to attend part of the workshop days to develop an understanding of the PD model being used.

- *Liaison with participating schools:* Assist with the publicising of the project and recruiting schools to apply. Once schools were selected they acted as a point of contact, for example, communications about workshop days and if there were any issues for participating schools they were the point of contact alongside the PD Leads for these to be raised.
- *Evaluation and learning lessons*.

The MESH were expected to contribute to the on-going evaluation of the project in order that lessons could be learned for future years.

Core Teachers (CT): Each TIME team consisted of a pair of teachers from each of the10 schools. The pairs were an experienced teacher with some departmental responsibility and a less experienced or non-specialist teacher of mathematics. The teachers were responsible for using curriculum materials with their classes leading developments in their own schools, supported by the expertise of the group leaders.

At inception core teachers' responsibilities were identified as

- participating in the TIME team workshops
- delivering project lessons: collaborating together to plan and teach the unit lessons to their KS3 classes
- further professional development activities: gathering and analysing evidence of impact; carrying out assessment interviews with pupils, and carrying out one joint lesson study per unit and presenting feedback at the next TIME team workshop
- sharing with the department: engaging and sharing developments with the rest of the department

3.4 Professional development activities and project timeline

The key essential aspects of the project from the perspective of the teachers involved were:

- attendance at 5 professional development events involving 6 days of CPE (TIME meetings)
- introduction to lesson study and a request to engage in at least one lesson study cycle
- the use of diagnostic assessment materials with pupils (Unit 0)

- access to and support to use a series of 'Lessons'[12]

Lessons were organised into three units, Unit 1, Unit 2 and Unit 3. For Units 1 and 2, 6 distinct lessons were developed, each developer contributing two in each unit. For Unit 3 each developer contributed a single lesson. The table below provides details and further information about project materials is available in Annex A.

Table 1 Curriculum materials

Unit 0	• Diagnostic and formative assessment tasks					
Unit 1	1a. Parts of a shape	1b. Pieces of a cake	1c. Fair Shares	1d. Our survey said	1e. Ordering and equivalence	1f. Milkshakes
Unit 2	2a. Contexts and the bare model	2b. Percentages on the bar model	2c. Identifying proportional scenarios	2d. Directly or indirectly proportional	2e. using the DNL to explore relations	2f. Using the DNL to solve ratio
Unit 3	3a &b Ratio tables		3c & d Using stories and diagrams to model multiplication and division		3 e & f Exploring multiplicative structures	

The project was planned to start in May 2013 with recruitment of project schools completed by July 2013 (see Annex A). External evaluation procurement and the requirements of an RCT design meant that it was not possible to begin the project as intended in July 2013. Thus the intended timetable was slightly delayed. The table below gives the key milestones and activities in the project; a detailed project narrative is included in the project. Note that in the initial intended timetable, a number of departmental workshops were envisaged. During the inception phase the project focus on the Core Teachers' professional learning are further emphasised and so these were not a formalised aspect of the programme.

[12] In general it was agreed that a project lesson would generally involve more than one single hour timetabled lesson when taught.

Table 2 Project timeline

May 2013	Project inception. Recruitment of MESH schools, PD Leads, university experts and curriculum developers for the TIME teams Recruitment of external evaluators and agreement of NAG membership
June 2013	Recruitment of MESH schools, PD Leads, university experts and curriculum developers for the TIME teams completed Recruitment of project schools Agreement of NAG membership Core team develops programme for the project schools First round of recruitment
July 2013	Recruitment of external evaluators and NAG (1/2 day meeting) Appointment of NCETM project lead 2 day Core design team initial meeting (**CDT 1** July 25/26th) – project design
August 2013	Appointment of external evaluators
September 2013	School application documentation developed **CDT 2** – Sept 5th, Core team develops resources for formative assessment and first taught units National development team -**NDT 1** (2 day planning and development workshop: overview, formative assessment units, on-going evidence collection)
October 2013	Recruitment of schools completed Allocation of schools to control and intervention groups by RCT evaluators **TIME team workshop1** (2 day workshop in the 3 local venues: Understanding the issues, formative assessment units and collecting evidence) **CDT3** – unit 1 completion
November 2013	TIME team workshop 1 continued Formative assessment units - Unit 0 used with Y7/8/9 Classroom observation and evidence collection **NDT 2** (1 day planning and development workshop: Units 1 and 2)
December 2013	**TIME teams** (1 day workshops in 3 local venues: Unit 1) **CDT 4** unit 2
January 2014	First intervention units taught to Y7/Y8/Y9 Lesson study, classroom observation and evidence collection NAG (1/2 day meeting) **NDT 3** unit2
February 2014	**TIME teams** (1 day workshop in 3 local venues: Unit 2) Second units taught to Y7/Y8/Y9 Lesson study & classroom observation and evidence collection **CDT 5** unit 3
March 2014	**NDT 4** (1 day planning and development workshop: Unit 3, final evaluation, taking the project forward) **TIME teams** (1 day workshop in 3 local venues: Unit 3) Interim assessment by RCT evaluators
April/May 2014	Third units taught to Y7/Y8/Y9 Lesson study, Classroom observation and evidence collection **CDT 6**
June 2014	Post-test **TIME teams** (1 day workshop in the 3 local venues: evaluation & embedding)
July 2014	**NDT 5** (1 day workshop)

3.5 Recruitment narrative

Schools were recruited to the project in a number of ways. The NCETM advertised the project through email, in newsletters and on the web. In addition, the DfE advertised the project in its communication to schools. Following their recruitment, the MESH hosts also promoted the project through their networks. By July 2013, 110 schools had expressed an interest. The Centre for Education and Inclusion Research (CEIR) and the NCETM worked together to produce information material for schools on the MRP and its evaluation. The 110 schools were then sent full information about the project, the nature of the randomised control design and asked to supply information on relevant KS3 pupils and potential CTs. From case study visits it appears that in some schools senior leaders did not agree to teachers who wished to participate applying to do so.

Some 66 schools completed an application. Issues were identified with 34 of the returns, involving incomplete or ambiguous information. CEIR and NCETM worked together to resolve this with both organisations committing additional resource to ensure information was obtained. Four schools either withdrew or were excluded at this point, due to ineligibility.[13] This meant that 62 schools were identified as wishing and eligible to participate with 62 deemed eligible to participate.

Given the recruitment of 62 schools, 30 were allocated to the intervention and 32 to the control. In each area, 10 schools were randomly selected for the MRP intervention. This was done using a stratification scheme based on geography and school level (GCSE) attainment

After randomisation, in region B, it was identified, that one of the schools placed into the control group was in the same teaching school Alliance as the MESH host for that region, with the host also supplying the PD leads. The importance of avoiding release of project materials was emphasised to the MESH and PD leads and the control school continued to participate in the project.

In region C, an issue was identified that both of the schools supplying PD leads had applied to the join the programme. At the point of randomisation, the information we had did not have details of PD leads, otherwise the two schools might have been excluded from the trial. Both schools were allocated to the intervention group. Given this, whilst there was some potential of 'increased treatment' we did not judge this would put the overall trial at risk and so the two schools remained in the trial.

These two situations could have been avoided if there had been an earlier appointment of the external evaluator during the design phase. Advice could have

---

[13] Reasons for exclusion were: nominating teaching assistants or internal consultants without KS3 class teaching responsibility or a new academy without enough teachers, teaching KS3 so that one school could only nominate one core teacher.

been given about the implications of an RCT for issues of independence of schools and the implications of this for recruitment.

The nature of recruitment - having both a national and a local component - meant that the regional networks combined being geographically spread with some clustering around the MESH hosts' locations. A school with the furthest travel distance from a MESH were approximately 4 hours' drive away, with others very close.

The issue of two teachers being out for 6 days appears to have also effected recruitment, one HoD stated that she knew other schools that wanted to apply but the headteacher blocked them being involved. This was reported as being a particular issue for schools that were in special measures or lower attaining.

4. Evaluation Methodology

4.1 Overview of the trial and evaluation methodology

4.1.1 Overview of the methodology

A 3-level Clustered Randomised Control Trial (CRT) was used to assess the potential causal impact of the MRP on the development of mathematical skills and understanding, including multiplicative reasoning. Three parallel clustered RCTs in Years 7, 8 & 9 in order to examine impact at all levels of key Stage 3. Key Stage 2 scores were used as a pre-test measure and the outcome measure was the GL Progress in Mathematics Tests. The RCT was supported by a process evaluation that involved national, regional and school level data collection through surveys, interviews and documentary analysis.

4.1.2 Evaluation timeline

The table below provides a summary of the timeline of evaluation activities

Table 3 Evaluation timeline

Time of year	Activity
August-September 2013	Recruitment of schools
October 2013	Data gathered from schools to confirm Core Teachers and classes to be taught by the Core Teachers. Randomisation of schools into control and intervention groups. All schools contacted with initial information about the project. Schools distribute parental opt-out consent letters.
December 2013	Telephone interviews with National and PD leads
February 2014	Evaluation visit to NDT. Pupil numbers to be tested confirmed with schools and initial testing information sent to schools.
March 2014	Evaluation visits to TIME events begin. Schools provide information on any movement of pupils between classes, any pupils no longer at the school and any details on unexpected circumstances (e.g. CT unable to attend NCETM events etc).
May 2014	NPD data received from DfE. This included pre-test data (Key Stage 2 finely graded mathematics point score), date of birth information (used to calculate the age-related post-test score), pupil eligibility for free school meals and gender.
June 2014	PiM test undertaken in all schools in sample to gain the post-test score (tests took place on the 4th June or within that week). Case study visits to 9 schools begin

June/July 2014	PiM tests marked by GL assessment. Test results distributed to schools in July. Evaluation visit to final TIME event. Case study visits to 9 schools. Exit survey of intervention and control schools
August 2014	Datasets brought together and initial analysis undertaken
September/October 2014	Final analysis and report written

4.2 Evaluation questions

Research questions were formulated in response to the aims and aspects of the evaluation set out by the DfE.

Aim A: to test the impact of the programme

Research Question (RQ) 1: What is the impact of the programme on pupil outcomes on both general mathematical attainment as measured by GL assessment Progress in Mathematics tests and on those items in the GL PIM associated specifically with multiplicative reasoning?

RQ2: What are the impacts (if any) on: pupils' relationships to mathematics; teacher beliefs and practice including on lesson planning; teacher knowledge of multiplicative reasoning pedagogy; capacity of core teachers' to lead professional development?

Aim B: to assess why/how the programme worked or didn't work

RQ3: How is the programme conducted and if this differs from the planned programme in what way and why?

RQ4: What are the views of teachers/development teams on the programme including its effectiveness?

RQ5: If/how was the programme effective and what lessons can be learnt for scalability?

RQ6: Are there any patterns of differences in effectiveness for particular groups of pupils, teachers, schools, or across the three TIME teams)?

Aim C: to assess whether the programme is cost effective

RQ7: What was the delivered through the programme including: activities; quality and quantity of professional development; the reach (teachers and pupils) including beyond those directly involved impacts on organisational capacity?

RQ8: Was the programme cost effective?

To meet these aims we developed a methodology (discussed in detail below) that comprised a Randomised Control Trial (RCT) and a Process Evaluation (PE) conducted in parallel and that was followed by an integrative analysis that related outcomes from the RCT and PE.

4.3 Ethics and consent

4.3.1 School and teacher consent for participation in the trial

Prior to randomisation, schools were sent information about the scope of the trial (including the possibility of being approached to host a school visit after the summer testing should their school be allocated to the intervention group). Schools were requested to make arrangements to distribute an information and consent letter (created by SHU) to the parents/guardians of all pupils taught by the school's nominated core teachers. Thereafter, headteachers were required to sign and return a consent form that confirmed the following:

- the information and consent letter had been sent to parents
- an agreement for the school to inform NCETM/SHU of any parents that returned opt-out consent forms (see 4.3.2)
- an agreement to take all reasonable steps to ensure the required sample of pupils took the GL Assessment PIM test in June 2014
- an agreement for the school to take part in the trial
- an understanding that participation was voluntary and that their school was free to withdraw from the trial at any stage
-

- The consent form returned by the headteacher also included the signed agreement to participate in trial by the two nominated core teachers.

4.3.2 Parental opt out consent

As outlined above, the lead core teacher in the school was asked to distribute parental opt-out consent letters to the parents of those pupils involved in the study. This was done in the autumn term 2013. Parents returned opt-out slips to the core teachers who then informed SHU of any pupils who had been opted out of the study. Parents could opt their child out of the study at any point during the project. In addition, some teachers also felt that the test would cause undue stress to some of their pupils; therefore some teachers withdrew some of their pupils for ethical reasons. In one instance, a school withdrew an entire year group from the trial because of the close proximity of the GL test to their statistics GCSE.

4.3.3 NPD and data protection

NPD data was requested from DfE in spring 2014 and received in May 2014 (please see section 4.1.2 for further details). This data was shared with SHU under strict accordance with data security policies agreed between SHU and the NPD team at DfE.

4.3.4 School visit participation

As outlined in 4.3.1 school headteachers and core teachers entered into the study were aware of the possibility that their school might be approached to host a school visit. Following the GL Assessment test in June, an email with an information sheet attached outlining the proposed structure of the visits (see 6.2.3) was sent to the lead CT and HoD for mathematics requesting their school to host a visit.

All researchers undertaking school visits were highly experienced at working within schools and were in receipt of Disclosure Barring Service (DBS) documentation. Schools were requested to arrange pupil focus groups for the school visit (ensuring no pupils whose parents had provided opt-out consent forms were present) and were given the option to have a member of school staff present should they want to. All participant information clearly outlined the purpose of the project to an age appropriate level, the intention to audio record interviews (subject to agreement from all participants present) and made clear how data was intended to be used. All focus group participants and interviewees were requested to provide audio consent prior to the commencement of interviewing.

4.3.5 Reporting

In order to comply with the terms under which schools, parents and individual teachers consented to the project, a series of precautions have been undertaken to avoid identification of any of the aforementioned within the report. The three TIME team regions have been allocated a regional code A, B, C, each school a unique number 1-60 and staff/pupils fully anonymised for example CT1 represents the lead core teacher. In addition, care has been taken to avoid any quotations or narratives that might be likely to make any individual, region or school particularly identifiable.

5. Randomised control trial

5.1 Trial Design

The 3-level Clustered Randomised Control Trial took place between October 2013 and June 2014 with multilevel analyses taking place between July and October 2014. Prior to randomisation (see section 5.4 below), participating schools were required to provide details of two named core mathematics teachers who taught maths to Y7, Y8 and Y9 classes. In addition to the named core mathematics teachers, schools also provided a list of pupils taught by these named teachers (names, dates of births and Unique Pupil Numbers). It was important to collect these details before randomisation to ensure as best as possible, an 'intention to treat' standard.

A total of 62 schools provided the required details about two core maths teachers and the Y7 to Y9 pupils they taught. 30 schools were then randomly selected to become intervention schools. In these 30 intervention schools, the two named core mathematics teachers received the Multiplicative Reasoning Programme via the PD events and supporting materials. The remaining 32 schools became the 'business as usual' control group. Immediately following randomisation, 2 control schools withdrew from the trial, leaving 30 schools (with 60 core maths teachers and their Y7 to Y9 pupils) within the control group. Baseline details on all pupils within the 60 participating schools were extracted from the National Pupil Database[14].

The trial was publically registered through the Current Controlled Trials website[15] and was given a unique reference of ISRCTN63650913 on April 24th 2014. This follows best practice CONSORT guidelines[16].

5.2 Intervention and control activity

Details of the MRP intervention are given in section 3 above. The intervention had both a professional development and curriculum development aspect. Figure 2 provides a model of a theory of change, indicating how changes in pupil learning might occur.

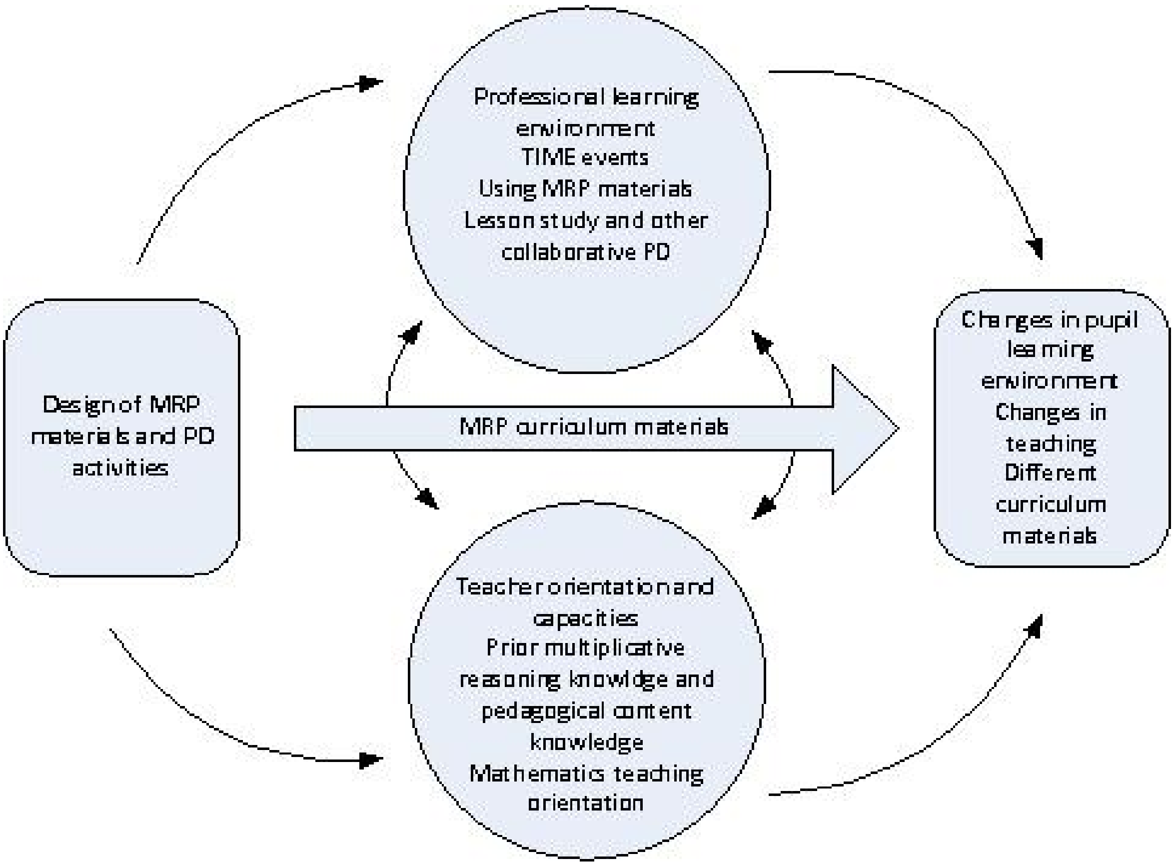
---

[14] The National Pupil Database (NPD). For this trial, attainment at KS2 maths, gender, FSM and SEN status was extracted from the NPD using the pupil names, dates of births and unique pupil numbers provided by schools prior to randomisation.

[15] For the public registration of this trial, please see control trials website.

[16] See consort website.

Figure 2 The MRP theory of change



The theoretical basis for the project supposed that improvements in pupils' MR capacity might lead to a more general improvement in mathematical attainment.

The control schools were anticipated to continue on a 'business as usual' basis.

5.3 Sample size

At the design stage, a power analysis was conducted using the Optimal Design Software[17] to explore the minimum detectable effect size (MDES) for the proposed 3-level clustered RCT design[18]. This power analysis assumed the following:

[17] Optimal Design power analyses software and associated manual are freely available.

[18] Three-level clustered randomised control trial with pupils clustered into classes clustered into schools.

A statistical power of 0.8[19] (or 80%)

- number of schools = 60;
- average classes per year group = 2 (6 classes for the combined Y7-Y9)
- average pupils per class = 25
- 5 percent of variation in the outcome will be between classrooms (class level intra cluster correlation coefficient).
- 10 percent of variation in the outcome to be between schools (school level intra cluster correlation coefficient)
- the pre-test covariate (KS3 maths score) will be used and this is assumed to statistically account for (or explain - using the Pearson's r^2 statistic) some of the variation in the outcome measure(s)

- It is also assumed that the statistical association between the pre and post test measures will diminish with time. For example, the correlation between KS2 attainment and PIM12 will be greater than KS2 v PIM13 which in turn will be greater than KS2 v PIM14. The Pearson's R^2 estimates used will be 0.4 (40%) for the Year 7 sample (KS2 v PIM13); 0.3 for the Year 8 sample and 0.2 for the Year 9 sample. For the combined Y7 to Y9 sample, the Pearson's R^2 estimate was 0.3 (30%).

From these analyses, at the point of randomisation the following predicted minimum detectable effect sizes (MDES) were calculated - and are shown in Table 4 below.

[19] This means that the CRT design is estimated as being able to detect a specific effect size with 80% confidence of not making a Type II error (or false negative) - whilst simultaneously having 95% confidence of not making a Type I error (false positive). A false negative (Type II error) is when an effect really does exist but is found to not be statistically significant in the analysis - and so the wrong conclusion of 'no effect' is made. A false positive (Type I) error is when an effect does NOT really exist but is found to be statistically significant in the analysis - and so the wrong conclusion of 'genuine effect' is made. Type I and II errors are linked so that reducing one leads to inflating the other - power analysis tends to focus on keeping the probability of making these error as small as possible by improving the design and/or increasing sample size.

Table 4 PREDICTED Minimum Detectable Effect Size (MDES) statistics

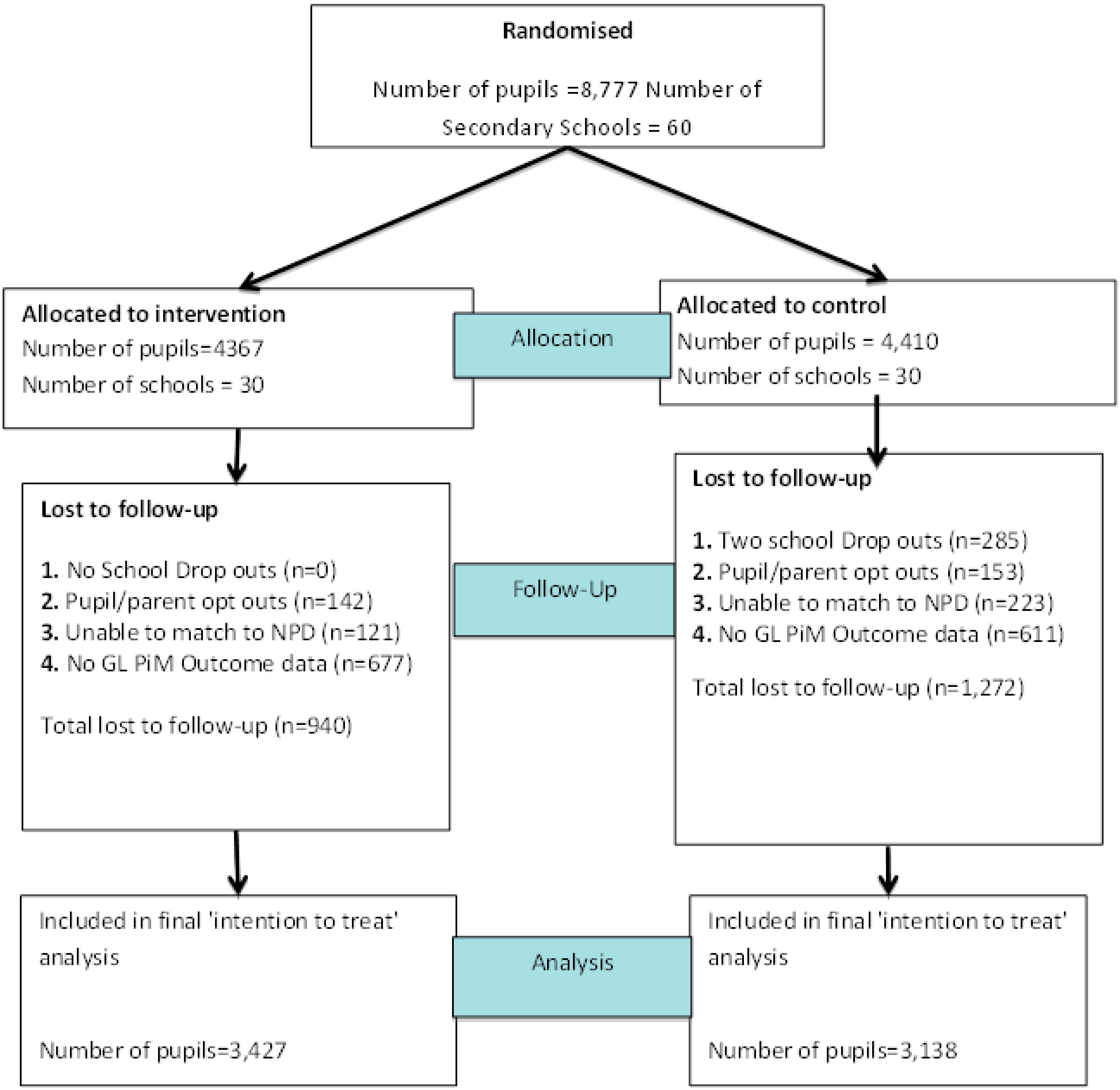
Sample size estimated as 3,000 pupils per year group, 9,000 pupils for the combined Y7 to Y9 sample.

	Y7	Y8	Y9	Y7 to Y9
Number of schools = 60, Ave classes per year group = 2 Ave class size =25	d/h[20]=0.24	0.25	0.26	0.22

At the analysis stage, the power analyses were re-done incorporating the actual rather than estimated / assumed detail. A statistical power of 0.8 (80%) was still assumed.

[20] Cohens d or Hedges g effect size statistics. Cohens d is the mean difference effect size based on units of standard deviations for the outcome (PiM) variable. Hedges g is a very similar effect size statistic to Cohens d and is the mean difference also in units of standard deviations for the outcome. The difference between d and g is that hedges g draws on the standard deviations for both intervention and control group samples whilst Cohens d uses the overall standard deviations once the control and intervention samples are combined.

Figure 3 Consort diagram for the trial



The final achieved sample with complete baseline and outcome data was 6,565 pupils across 349 Y7, Y8 and Y9 classes in 58 secondary schools. The following details were updated using the achieved rather than estimated sample size:

- final number of schools was 58 for Y7 to Y9 sample, between 52 (Y9) and 56 (Y7)
- average classes per year group = 2 (Y7 to Y9 combined and within each year group);
- average pupils per class = 19 (Y7 to Y9 combined and within each year group)

- class level ICC - 52% for the combined Y7 to Y9 sample[21], 42% for Y7; 47% for Y8 and 70% for Y9
- school level ICC - 18% for the combined Y7 to Y9 sample[22], 21% for Y7; 26% for Y8 and only 2% for Y9

The statistical association between the pre and post-test measures was empirically shown to diminish with time[23]. For example, the correlation between KS2 attainment and PIM12 was observed to be greater than KS2 v PIM13 which in turn was observed to be greater than KS2 v PIM14. The Pearson's R^2 correlations were observed to be 0.70 (70%) for the Year 7 sample (KS2 v PIM13); 0.69 for the Year 8 sample and 0.59 for the Year 9 sample. For the combined Y7 to Y9 sample, the Pearson's R^2 statistic was observed to be 0.66 (66%).

The average class size was lower than anticipated (19 rather than 25) but this will have a limited impact on the statistical power and MDES calculations. The structure of variation is somewhat different to what was anticipated at the design stage - with around 50% of the total variation in the (overall) PiM outcome variable located at the class-level and 20% located at the school level. This leaves 30% of the variation at the individual pupil level.

The proportion of variation located at higher levels within a multilevel design is known as the Intra Class Correlation Coefficient (ICC). The difference in observed ICC statistics with what was anticipated will impact on the MDES - and these are shown in Table 5 below. The notable class level ICC (of 50%) is an indication of how 'similar' pupils are (in terms of attainment) within each maths class - in other words, this is an ability grouping effect. Ability grouping brings together pupils with similar levels of mathematics into 'sets' or 'streams'. It is interesting to observe how this class-level ICC is much stronger for the Y9 sample (70% of the variation located at the class level) compared with Y8 (47%) and Y7 (42%). This may reflect a movement to universal or near universal setting by ability group between Y7 and Y9.

---

[21] The proportion of variation that is clustered at the class-level was found to be notably greater than anticipated. This is likely to reflect the widespread use of ability group setting and streaming within KS3 (Y7 to Y9) maths classes - clustering pupils of a similar attainment together into classes. The models picked up how classes consisted of 'similar' pupils in terms of attainment - and 52% of the variation in the outcome was found at this class level for the combined Y7 to Y9 sample.

[22] The proportion of variation that is clustered at the class-level was found to be notably greater than anticipated. This is likely to reflect the widespread use of ability group setting and streaming within KS3 (Y7 to Y9) maths classes - clustering pupils of a similar attainment together into classes. The models picked up how classes consisted of 'similar' pupils in terms of attainment - and 52% of the variation in the outcome was found at this class level for the combined Y7 to Y9 sample.

[23] See Annex E

This variance structure will impact on the achieved MDES, as illustrated in Table 5 below.

Table 5 ACHIEVED Minimum Detectable Effect Size (MDES) statistics

Achieved sample size of 6,656 pupils for the combined Y7 to Y9 sample.

	Y7	Y8	Y9	Y7 to Y9
Actual number of schools	55	56	52	58
Average number of classes per school	2.3	2.0	2.1	6.0
Average number of pupils per class	18.7	18.9	18.9	18.8
R2 (KS2 v PiM)	0.70	0.69	0.59	0.66
ICC (class level)	0.42	0.47	0.70	0.52
ICC (school level)	0.21	0.26	0.02	0.18
Minimum Detectable Effect Size (MDES)	**d/h= 0.41**	**0.43**	**0.48**	**0.29**

Overall, the achieved MDES for the combined Y7 to Y9 sample was 0.29 standard deviations rather than the estimated 0.22 standard deviations. According to the EEF (Higgins et al., 2013), both the estimated and achieved MDES fall within the 'moderate' description and can be roughly converted into between 3 and 4 months progress in school.

5.4 Randomisation

Within each of the three regions, randomisation was done by stratifying on attainment - and using the national average of 59% attaining 5+ A*-C (taken from the 2011/12 school performance tables)[24].

Schools with a greater proportion of pupils attaining this level (i.e. 60% or higher) were placed in the 'above average' group whilst other schools were placed in a second 'average or below average' attaining group[25].

---

[24] School performance tables

[25] This was done for all schools with GCSE attainment details. There was an instance when this data was not available and so judgement was used. In SOLENT, one Middle School was placed in the higher attaining group based on FSM and IDACI profile.

The result is illustrated in Table 6 below. The first number represents the total number of schools in each strata whilst the second one in brackets indicates the number of schools that were randomly assigned to become intervention schools.

Table 6 Randomisation

	Lower Attaining	Higher Attaining
Region 1	6 (3)	15 (7)
Region 2	4 (2)	15 (8)
Region 3	4 (2)	18 (8)
Regions 1 to 3 combined	14 (7)	48 (23)

Randomisation was done using the SPSS statistical software package. The SYNTAX and guidance notes can be found in Annex E. The SYNTAX and guidance notes were created by one researcher[26] and the actual randomisation process was conducted by a different researcher[27]. Randomisation took place on October 7th 2013.

Immediately following randomisation, two schools withdrew from the trial after discovering that they were placed into the control group. The result was a final sample of 60 secondary schools - half of which were randomly selected to receive the MRP and the remaining half formed a 'business as usual' control group.

5.5 Outcome measures

5.5.1 Rationale

The test instruments were GL Assessment Progress in Mathematics (PIM) 12 (for Y7), 13 (for Y8) and 14 (for Y9). These are standardised tests of general mathematical capability that are age specific. The reasons for using a general mathematics test were:

- proven success of PIM tests in a previous RCT of Every Child Counts[28]

---

[26] Sean Demack wrote the SPSS SYNTAX used within the stratified random allocation.

[27] Anna Stevens ran the SPSS SYNTAX to randomly allocated schools into the control and MRP intervention groups - See Annex E for a copy of this SYNTAX

[28] Torgerson, et al. C. (2013) Every Child Counts: testing policy effectiveness using a randomised controlled trial, designed, conducted and reported to CONSORT standards. *Research in Mathematics Education* 15(2): 141-153

- cost effectiveness and efficiency for large groups being tested of using an existing test
- independent marking by a reputable supplier
- the importance of assessing the impact on general mathematical ability as well as MR (to examine if there is a curriculum narrowing effect)
- it allowed comparison against national test data - since the test is widely used - potential supporting the cost effectiveness analysis if impact had been found by providing a measure of acceleration in attainment against national data

In addition the multiplicative reasoning elements of the test could be isolated by expert analysis of those test items related to multiplicative reasoning (undertaken Professor Iszák).

5.5.2 Pre-test

Attainment in Key Stage 2 mathematics was used as a pre-test or prior attainment measure within the impact analyses. See Section 7.2 for a summary of KS2 maths attainment, how KS2 maths attainment for the intervention and control group samples compare, the impact of sample attrition and the correlations between KS2 maths attainment and the PiM outcome measures.

5.5.3. PIM as post-test

The main post-test measure used was the standardised age score (SAS) calculated from the results of the PiM test. This measure is based on the underlying raw score so that it takes into account the age of pupils and allows the score to be compared to the national average. In addition to overall attainment in mathematics captured by the PiM SAS, three PiM subscales were used that focused on items within the PiM test that were identified as being related to multiplicative reasoning. Overall and across the three year groups, a total of 12 outcome variables were used as shown below in Table 7.

Table 7 Summarises the GL PiM outcome variables by year group

	PiM standardised age score	MR related but not necessarily related to the MRP	MR related and weakly or strongly connected to the MRP.	MR related and strongly connected to the MRP
Y7	PiM12	PiM12_MR	PiM12_MRW	
Y8	PiM13	PiM13_MR	PiM13_MRW	PiM13_MRS
Y9	PiM14	PiM14_MR	PiM14_MRW	PiM14_MRS
Y7 to Y9 combined	PiM12to14			

Essentially, for the main impact analyses there are three key outcomes (PiM12, PiM13 and PiM14) with one additional combined outcome (PiM12to14). In addition to this, subscales of the PiM test scores were created focused in on PiM test items relating specifically to multiplicative reasoning. The derivation of these subscales drew on the work of Andrew Izsák in consultation with NCETM. In all, the analyses involved a total of 12 PiM outcome variables.

Table 8 summarises how the PiM items were classified into the following four categories:

- not related to multiplicative reasoning at all
- related to multiplicative reasoning but does not connect directly with topics covered in model lessons
- related to multiplicative reasoning and weakly connected with topics covered in model lessons
- related to multiplicative reasoning and strongly connected with topics covered in model lessons

Using these categorisations and in consultation with NCETM, three subscale measures were derived for each of the PiM outcome measures.

The first subscale, focuses on items that have any relationship with multiplicative reasoning (PiM12_MR to PiM14_MR in Table 1). The second subscale focuses in on items that have a relationship with multiplicative reasoning *and* were identified as being related (weakly or strongly) to topics covered in the MRP model lessons by Professor Izsák (PiM12_MRW to PiM14_MRW in Table 1). The final subscale focuses in further on items that have a relationship with multiplicative reasoning *and* were identified as being strongly related to topics covered in the MRP model lessons by Professor Izsák. For these final subscales, only items on PiM13 and PiM14 were identified as being strongly related to the topic covered and so these subscales were created just within the Y8 (PiM13_MRS) and Y9 (PiM14_MRS) subsamples.

Table 8 Summary of categorisation of the PiM test items

	PiM12	PiM13	PiM14
Not related	Q1a, Q2, Q5a, Q6a, Q6b, Q6c, Q9c. Q10a, Q11a, Q11b, Q11c, Q12a, Q12b, Q13a, Q16a, Q16b, Q18, Q19b, Q20, Q22, Q23 21	Q6a, Q6b, Q8a, Q8b, Q10a, Q10b, Q11, Q12a, Q12b, Q13, Q14, Q16a, Q17, Q26a, Q26b, Q30a 16	Q1a, Q9, Q11a, Q11b, Q12, Q14, Q17a, Q17b, Q19a, Q19b, Q20a, Q20b, Q21b, Q24a, Q24b 15
related but unconnected	Q1b, Q3a, Q3b, Q4a, Q4b, Q7, Q8a, Q8b, Q9a, Q9b, Q14a, Q14b, Q14c, Q15a, Q15b, Q17a, Q17b, Q19a, Q26a, Q26b, Q28 21	Q3, Q4a, Q4b, Q5, Q16b, Q19, Q21, Q22a, Q22b, Q23, Q24a, Q27a, Q27b, Q28a, Q28b, Q28c, Q30b 17	Q2, Q3, Q4a, Q5, Q7a, Q7b, Q8c, Q13a, Q13b, Q13c, Q15a, Q15b, Q18, Q21a, Q22a, Q22b, Q23, Q25, Q26a, Q26b, Q26c, Q28a, Q28b, Q29a, Q29b, Q29c 26
weakly connected	Q5b, Q10b, Q13b, Q21, Q24, Q25a, Q25b, Q27 8	Q1, Q2, Q9a, Q9b, Q9c, Q9d, Q18, Q20, Q24b 9	Q1b, Q6, Q8a, Q8b, Q27 5
strongly connected		Q7a, Q7b, Q15, Q25, Q29a, Q29b, Q29c, Q29d 9	Q4b, Q10a, Q10b, Q16 4

See Section 7.2 for a summary of the PiM outcome measures, how attainment in PiM for the intervention and control group samples compare and the correlations between attainment in PiM and KS2 maths.

5.5.4 Scoring and marking

The majority of tests took place in school on the 4th of June or within that week. Invigilators then collected in the papers and left them at the school to be collected by GL Assessment for marking. The papers were received by GL Assessment shortly after the testing period, and marked over a 2 week period. Around 200 papers were marked in August due to a delay in obtaining date of birth information (used to calculate the standardised age score). Individual school reports were produced and distributed to schools before the end of the summer term. A full dataset of results in excel format was returned to Sheffield Hallam University from GL Assessment and integrated with the NPD and other project information (such as pupil movement between classes) using an anonymised unique identifier.

5.6 Other measures used within the analyses

5.6.1 School Context

Randomisation occurred at the school level within the clustered randomised control trial (CRT); 30 schools were randomly allocated to receive the MRP and the remaining 30 schools formed a 'business as usual' control group.

A school level comparison of the MRP intervention and control groups was conducted (see section 7.2 below) and some imbalance was observed. To try to statistically take account of this school-level baseline imbalance, the following school-level variables were included into the impact analyses; school GCSE attainment; admissions policy, IDACI score, percentage of pupils classed as FSM[29] and school OFSTED rating.

5.6.2 Other pupil-level measures

In addition to attainment in KS2 mathematics, the following pupil-level variables were included into the impact analyses; pupil gender; whether a pupil is classed as 'FSM' and whether a pupil had a Special Educational Need (SEN).

---

[29] The FSM measure identifies young people who *are* eligible and claiming free school meals from young people who are *not* eligible or eligible and not claiming them. Although this measure is widely used as a proxy for the socio-economic status of young people in England, it should be noted that there is a known undercount of FSM claimants that is estimated at approximately 200,000 (or 2%) of all 4-15 years olds in England (Iniesta-Martinez and Evans, 2012). Whilst FSM is a rather simple measure of socio-economic status and has this problem of inaccuracy, it is readily available and until better socio-economic detail is collected, remains likely to be the main tool for taking socio-economics into account in educational research in England. The measure is used at the pupil to identify individuals who are 'FSM' and the relative concentration of pupils classed as FSM

See Section 7.2 for a summary of these other pupil level measures.

5.7 Conduct of the trial

At the point of submitting our proposal, SHU entered into an agreement with GL Assessment to be responsible for the provision, dissemination and marking of the post intervention test. SHU also secured an agreement in principle from a well-known national recruitment company to provide staff to invigilate the GL test in June.

As outlined in Section 6.2.4.1, prior to randomisation SHU worked closely with NCETM to ensure key data pertaining to each proposed core teacher was obtained from across all interested schools that applying to be part of the trial. On the basis of the information received decisions were made about which schools remained eligible for randomisation and those that needed to be excluded. Following randomisation, NCETM then got in touch with all schools informing them whether they had been randomly allocated to either control or intervention. In keeping with CONSORT guidelines arrangements were made to publicly register the RCT.

Key points of formal communication with schools

October 2013: Email with attached letter (orientated towards whether a control or intervention school) sent to the headteacher and core teachers of each school involved in the project that formally welcomed them to the trial. Included within the letter was a provisional timetable for future correspondence with an evaluation team action and corresponding school based action against each key milestone. A named member of the CEIR evaluation team was also identified as each school's primary contact for sending subsequent documentation and/or if they had any queries about any aspect of the trial.

December 2013: Email sent to the headteacher giving further details of what the final GL Assessment report following marking of the tests would look like; along with a request the name & contact details of the person they wished SHU to correspond with regarding test arrangements.

March 2014: Email sent to either key contact (if provided) or two core teachers, with a request to thoroughly check an attached spreadsheet list of pupils scheduled to take the test in June and to update the spreadsheet making clear:

- any pupils they had received opt out consent forms for that remained on the list
- any pupils no longer at the school

- any pupils that had changed class at the point after the school was selected to take part in the research

The email also included was a provisional testing guidance documents outlining key requirements for the test along with a testing arrangements document that sought the school to provide the following information and return to SHU:

- confirmation the test could take place on June 4th
- test start time, venue for against each year group
- name and contact details of the examinations officer

April - May 2014: A series of email and telephone calls chasing outstanding information from the March request around pupil lists and testing arrangements.

May 2014 - Email sent to school with the following key documentation

- *School testing guidance document*: This provided a detailed account of the timeline until the testing date. It also outlined the responsibilities of the school, invigilators and the school.
- *Agreed testing arrangements*: This details the arrangements made with the school.
- *Excel spreadsheet*: Based on the information provided by the school. This reflects back the list of pupils we wanted to test
- *Frequently asked questions.*
- *Invigilator script*: for reference. This is what the external invigilator used to introduce the test.

-
- **May 28th 2014**: Email sent to confirm the courier delivery of the tests had been received.

August/September 2014: Email to schools providing passwords so schools can access their GL Assessment account of marked pupil's tests.

Invigilation arrangements provision

By January it had become apparent that the national recruitment company named on the bid would no longer be able to provide invigilators for the project. Following discussions with DfE, SHU made alternative arrangements with four other regionally based specialist education recruitment agencies to provide invigilators to fill the void. Each recruitment agency was provided with the following information:

- invigilator test guidance document: feedback Fform
- invigilator script –FAQ sheet

Regrettably at an advanced stage in the process one recruitment agency withdrew due to key staff leaving; this required SHU to bring in a national company at a late replacement to lead on provision for schools in region B and to also fill gaps where other regionally based recruitment agencies areas could not guarantee invigilators.

6. Process evaluation methodology

6.1 Approach/overview

The process evaluation (PE) was conducted in accordance with EEF guidelines for PEs occurring concurrent with RCTs as well as informed by the CEIR's extensive experience of PE including for national DfE initiatives. Data was collected at national, regional, school and teacher level and the components that form each are summarised below. [30]

6.2 Data collection

6.2.1 National level data collection

Telephone interviews and ongoing communication with NCETM

Telephone interviews were undertaken with NCETM secondary directors and the NCETM project lead at the beginning of the project in order to derive a more thorough grasp of the strategic direction of the programme and to clarify team structures and remits. There was also regular and ongoing verbal dialogue and email communication with the NCETM project administrator and project manager; documentation such as team structures, intended features of the programme, meeting schedules and timelines were all made freely available. SHU worked particularly closely with the NCETM administrator and project manager around the onset of the project (Autumn 2013) - in order to co-design appropriate documentation/materials for project schools. Meetings between the SHU evaluation team and the NCETM took place as needed and data were collected in this way.

NDT observation

The NDT meeting in February was attended in order to directly observe the process undertaken when planning PD resources, lesson materials and TIME team events etc.; as well as allowing space for informal discussions with the programme director and other key individuals to help ensure accurate understanding of the programme.

---

[30] Evaluation design and analysis was informed by: Boylan, M., Maxwell, B,, Coldwell M., & Y Jordan, J. (2014). Understanding professional learning: troubling concepts. Paper presented at *International Professional Development Association Conference,* November, 2014; Coldwell, M. and Simkins, T (2010) Level models of continuing professional development evaluation: a grounded review and critique. *Professional Development in Education* 37 (1); Clarke, D, & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education,* 18(8), 947-967. van Driel, JH.,et al. . (2012) Current trends and missing links in studies on teacher professional development in science education: a review of design features and quality of research. *Studies in Science Education,* 48(2), 129-160; Yin, R. K. (1994). Case study research: Design and methods (2nd ed.). Newbury Park, CA: Sage Publications.

In addition NCETM documentation and their internal evaluation data was also analysed.

6.2.2 Regional level data collection

Telephone interviews with PD leads

Telephone interviews were undertaken with PD leads attached to each of the three regions to determine their view on the project overall, to understand how previous NDT meetings operated, how the first/second TIME team meetings had gone and more broadly to gauge whether there were any general or specific concerns about the RCT picked up from the teachers.

Observation of Time team events

SHU researchers attended a TIME team workshop at each region - these visits were originally earmarked for December but at the request of NCETM postponed until February so the programme had more time to become established (due to a clash of appointments the researcher nominated to region B, needed to visit the subsequent March event instead). Region B was also visited in June (the final Time event) in order to gain a perspective on the NCETM's approach to evaluation. Overall, these visits helped the team to gain a deeper understanding of how materials were presented and engaged with, and to help determine if there were any fundamental differences in approach across the three TIME team regions.

TIME team event attendance data

Attendance data was requested from each of the TIME team events which helped inform judgements on levels of engagement across the intervention schools and also fed into the cost benefit analysis. Attendance data is given in Section 8.5.1.

TIME team paper surveys

SHU undertook short data capture surveys at three of the TIME team events to ascertain engagement and use of materials. It was not originally planned to do these short surveys but the decision was taken to do so in order to get information on material throughout the course of the project and to reduce the burden of a long survey at the end of the programme.

The surveys themselves were very contained and straightforward to complete (a double sided single A4 sheet of paper). Teachers were required to simply note approximately how many lessons they had dedicated to each unit for each class. The lack of requested detail was a deliberate compromise balanced against wanting to minimise additional burden on teachers, difficulties encountered with receiving information back from schools elsewhere and trying to ensure as full a response as possible. The surveys were distributed at the end of each of the regional time team meetings and then collected in by the NCETM project lead or sent directly back to SHU by the PD leads. Towards the end of the summer term, SHU emailed out a bespoke electronic version of the surveys to all intervention participants that reflected any outstanding surveys not received. The data from the surveys provides a proxy measure as to the extent to which core teachers have engaged with the materials. The table below outlines the number of core-teachers that completed the TIME team surveys.

Table 9 Number of respondents

TIME Team Meeting	Number of respondents*
February	48
March/April	44
June	33

*This may be an under estimate the number of teachers who returned surveys as some pairs completed one survey.

6.2.3 School level data collection

Baseline data capture

During the recruitment phase we retrieved data on school characteristics from publicly available sources to support the randomisations process.

School visits

One day visits to 9 schools (3 from each TIME team region) were arranged and undertaken following the June 4th test.

1.1.1.1.1 *Original approach to sampling for school visits*

The original sampling matrix sought to achieve a sample of 9 schools and identified 12 schools (4 schools for each TIME team - with one acting as a reserve) taking into account the following key features:

- *Unique aspects*: Two schools were purposively identified because they possessed unique features that were thought appropriate to investigate further. One where the PD lead also taught within the school and another where the classes initially identified and those taught were very different.
- *Level of engagement with the project*: Broadly sought to be sensitive to what appeared to be the extent of engagement with the programme, i.e. levels of attendance at TIME events and reported use of the materials.

- *Socio-economic range*: Attempted to include a range of schools across socio-economic contexts to represent the range of schools involved in the programme. For example, 25% of intervention schools were below the national average 5 GCSE A-Cs, so 2 of the 9 schools identified in the original sample were earmarked to be in that category.

The 12 schools originally identified were initially emailed details about the case study visits, which were followed up with telephone calls in order to determine willingness to be involved/to make practical arrangements for a visit.

Unfortunately, there was quite a limited response from the original 12 schools identified. The lack of receptiveness to host a visit is likely to have been influenced by a number of factors including:

- Lack of lead in time. Due to the relatively late timing of the test (June 4th) our focus was on working with schools/invigilation companies to ensure the test went as smoothly as possible. It was not thought to be appropriate for schools to be presented with additional requests for case study involvement during a pressurised exam period for schools in general and in relation to our specific test requirements for the project. This meant the 12 identified schools, were initially invited to host a case study from around the middle of June.
- Some of the schools identified were selected purposively because they appeared to have been relatively less engaged with project (i.e. attended less TIME team sessions and did not use the lessons as much). It is possible that the lack of engagement with the project in general made such schools by definition relatively less likely to respond positively to hosting a school visit.
- NCETM's parallel internal evaluation meant the pool of available schools was limited further as we did not feel it appropriate to approach schools that had already agreed to host an NCETM visit too (ultimately as the sampling criteria broadened in a small number of instances it was necessary to contact such schools in order to reach the numbers required).

1.1.1.1.2 *Revised sampling approach for school visits*

- The lack of response from originally identified schools (despite a number of attempts to get in contact) combined with the necessity to complete all case visits prior to the end of the Summer term determined the need to send out the case study invitation to all intervention schools. Of the original sample identified it was possible to visit 3 schools from the sample in region C, 1 in region B and C.
-
- Note that three of the schools were teaching schools and there is likely to be a sampling bias towards schools that had greater engagement and enthusiasm for the project.

Activities undertaken during school visits

School were asked to facilitate the following range of activities during visits:

- interviews with both of the core teachers (ideally separately)
- a interview with the head of mathematics or equivalent
- a interview with another teacher that was involved in lesson study or had also used project materials, or attended a CPD session delivered by a core teacher (if appropriate)

Additional activities requested but that were identified as being useful but not essential were:

- a interview with a senior leader
- a focus group with pupils that had recently used some of the materials from the project

Exactly what was possible at each school inevitably varied; for example at one school the other CT was not available because of a school field trip, at another the senior leaders were at an external meeting and thus not available. Below an outline of the composition of fieldwork activity at each of the nine case study schools is provided, along with basic school characteristics.

Table 10 Fieldwork activity by school

Region	School visited	FSM %	KS3 pupils	GCSE A-C %	Core teachers + other teachers interviewed	SLT/HT interview	HoD	Pupil FG's
Region A	1	4%	300	n/a	2	1	1	4
	2	3%	661	76%	2 + 1	0	1	0
	3	6%	750	71%	1	1	1	1
Region B	1	14%	200	59%	2 + 1	1	1	1
	2	7%	600	79%	2	1	0	1
	3	9%	480	69%	2	0	1	1
Region C	1	18%	428	68%	2 + 1 [trainee]	1 (PD lead)	1	2
	2	12%	650	66%	2	0	1	1
	3	11%	870	71%	1	0	1	1
Total					**19**	**5**	**8**	**12**

6.2.4 Teacher level data collection

Baseline data capture

Prior to randomisation, it was requested that NCETM include a series of questions in a data capture sheet for each proposed core teacher. Requested data included name, DfE teacher number (if known - to assist with NPD requests and to check qualified teacher status (QTS)), year of qualification as a teacher, detail of any initial teaching qualifications in mathematics, role in the department (if any) and proportion of teacher's subject teaching that was mathematics. This data formed part of the intention to treat analyses.

Exit surveys

Exit surveys were undertaken with both intervention and control schools.

Intervention schools

The intervention schools survey asked about engagement with PD, use of and views on materials and perspectives on project impact.

Overview of teacher level data collected from intervention teachers

Overall 42% of respondents completed all three surveys, 25% completed two of the surveys, 19% completed one survey and 13% did not complete any of the surveys.

Table 11 Number of surveys completed

Number of surveys completed	n	%
0	9	13
1	13	19
2	17	25
3	28	42
Total	**67**	**100**

Table 12 Number of respondents per TIME team meeting

TIME Team Meeting	Number of respondents
February	48
March/April	44
June	33

With regards to the exit survey, 21 teachers responded equating to around one third of the original sample. Around half of these respondents were case study schools.

Control schools

The comparison schools survey was a brief online survey to teachers in comparison schools that sought to help understand if there were any particular issues that might have influenced pupil performance at their school.

7. Impact Analysis

7.1 Methods of analysis

7.1.1 Multilevel Linear Regression Analysis

A multilevel regression approach was taken for the impact analysis within the MRP evaluation. Three levels were included into the models; pupils (level 1) clustered into maths classes (level 2) which are clustered into secondary schools (level 3). The outcome PiM variables are all treated as scale (ratio) level data and so a linear regression multilevel modelling approach was adopted.

7.1.2 Specifying the model

The main analyses will adopt an 'intention to treat' approach (see Section 7.3 and 7.4). This means that all available cases will be included regardless of fidelity to the MRP. A fidelity analysis was conducted (see Section 7.5) and following this, an analysis that adopted an 'on treatment' approach was undertaken.

The 'intention to treat' approach is regarded as the most valid one for identifying causal impact within an RCT design. This is because the selection of who was included was based solely on random methods (see Section 5.4). The only factor that undermines this is attrition and imbalance. However, the intention to treat approach ignores trial fidelity. Specifically, the analyses ignore whether a pupil moves away from the class they were in at the start of the trial, whether they were taught solely by the core maths teacher named prior to randomisation, whether this core maths teacher attended TIME events and reported to use any of the MRP materials. An on-treatment approach that focused in on a 72% subsample of the intervention group who did not move classes during the period of the trial, were taught solely by a named core maths teacher who attended at least 2 of the 4 TIME event days and reported to use some of the MRP material (see Section 7.6).

In addition to identifiers at the pupil, classroom and school levels, from the baseline analyses, the following explanatory variables were included into the model

…at the **pupil-level**:

- KS2 Maths Attainment (Fine Point Score) - centred around the mean
- FSM status (1=FSM, 0=not FSM)
- gender (1=Female, 0=Male)

- SEN (1=SEN; 0=not SEN)
- age in months (only used for PiM subscale outcomes)[31]

…at the **school-level**:

- group (1=Intervention; 0=control)
- school GCSE attainment (taken from the 2011/12 school census; 1=above average, 0=below average)
- school admissions policy (selective = 1, not selective=0)
- school IDACI score (based on postcode of school)
- school %FSM (taken from the 2011/12 school census)
- school OFSTED rating (1= grade 1 or 2, 0=grade 3 or 4).

The multilevel analyses were conducted using the STATA statistical software package and specifically the STATA MIXED procedure.

7.1.3 Stages of building the model

The multilevel models were constructed using four stages:

Stage 1: Empty Model

Stage 2: Group Only Model

Stage 3: Full Main Effects Model

Stage 4: Parsimonious Model

At stage 1, the **Empty Model** is used to assess the hierarchical structure of the data. The empty models will calculate the proportion of variation (in each of the outcome variables) that is attributed to each of the three hierarchical levels. This stage provides empirical details on the class and school level ICC (see Section 5.3).

The **Group Only Model** fits only the binary variable that identifies whether a participant was in the intervention (=1) or control (=0) group. This provides an indication of how different the PiM attainment for the intervention and control group is before taking other factors such as KS2 maths attainment into account.

The **Full Main Effects Model** fits all of the explanatory variables - i.e.

---

[31] The overall PiM attainment measure was a 'standardised age score' that used age in months of pupils to help determine the eventual attainment whilst the three MRP focused subscales were based on raw attainment without age standardisation. Therefore, for the three MRP subscales, pupil age in months was included into the group of pupil-level explanatory variables.

Pupil Level: FSM / Gender / SEN / KS2 Maths Attainment

Class Level: Class identifier

School Level Group / GCSE Attainment /selective school / IDACI / %FSM / OFSTED

The **Final Parsimonious Model** only includes explanatory variables that are observed to account for a statistically significant[32] proportion of the variation in the outcome variables. The one exception to this is the school-level group variable which will be included regardless of statistical significance. By excluding variables that are not contributing to the explanatory power of the model, the statistical power of the analyses is maximised - but in all cases, the findings of the full main effects model and final parsimonious will be compared before concluding whether a statistically significant difference between the intervention and control group samples is present.

The coefficient of the 'group' variable captures the statistical impact of the MRP programme on the PiM outcome measures. This coefficient is converted into an effect size statistic (hedges g) and 95% confidence intervals are provided. The approach for converting the coefficient into the hedges g effect size statistic can be found in Appendix E along with guidance notes on how to read the tables from the multilevel analyses.

7.2 Participation

7.2.1 Recruitment

At the point of randomisation (in October 2013 - the baseline) the MRP clustered RCT had 8,777 (potential) pupil participants who were taught in 418 Y7/8/9 classes by 120 core maths teachers in 60 secondary schools in England.

By the end of the trial (Sept 2014) there was complete data for 6,565 (75%) pupil participants in 349 Y7/8/9 classes taught by 116 core maths teachers in 58 secondary schools.

Whilst a 75% pupil-level response is good, it is worthwhile looking a little closer at where the 25% (n=2,212) were lost.

---

[32] A variable was retained within the model at the parsimonious stage if the p-value for the coefficient was observed to be 0.05 or lower.

Two (control) schools dropped out of the study during the course of the trial. This accounts for 13% (n=285) of the overall attrition at the pupil level. To comply with research ethics, pupils and their parents were able to opt out of participating in the trial by returning a form to their school - and this accounts for a further 13% (n= 295) of the attrition.

Sixteen percent (n=344) of the attrition related to missing baseline NPD detail for participants - this detail included KS2 Maths attainment and FSM status. KS2 maths attainment is an essential requirement in these analyses which focus on capturing mathematical progress of pupils between two points in time - KS2 maths to PiM12/13/14. Similarly, FSM status along with gender and SEN status will be included into the model as explanatory variables.

The remaining 58% (n=1,288) of the attrition relates to participants who did not take the GL PiM test in June 2014.

Table 13 below summarises response rates overall and for intervention and control groups separately across regions and year groups.

Table 13 Final sample sizes and response rates for the MRP clustered RCT

Overall and across year groups and regions.

	Control School Sample		Intervention School Sample		Total Sample	
	n=	% of baseline (response)	n=	% of baseline (response)	n=	% of baseline (response)
All Areas Combined						
Y7 to Y9 combined	3,138	71%	3,427	78%	6,565	75%
Y7	1,114	73%	1,274	79%	2,388	76%
Y8	1,067	71%	1,009	82%	2,076	76%
Y9	957	70%	1,144	75%	2,101	72%
Region 1						
Y7 to Y9 combined	1,047	66%	1,225	79%	2,272	72%
Y7	372	62%	474	75%	846	69%
Y8	345	64%	364	86%	709	74%
Y9	330	72%	387	80%	717	76%
Region 2						
Y7 to Y9 combined	827	72%	1,271	83%	2,098	78%
Y7	309	76%	487	84%	796	81%
Y8	228	70%	363	85%	591	79%
Y9	290	69%	421	80%	711	75%
Region 3						
Y7 to Y9 combined	1,264	76%	931	72%	2,195	74%
Y7	433	82%	313	78%	746	81%
Y8	494	77%	282	74%	776	76%
Y9	337	68%	336	65%	673	67%

Response was higher in the intervention sample (78%) compared with the control group (71%). This pattern was evident across the year groups but only in regions 1 and 2, in region 3 the response was higher for the control group (76%) compared with the intervention group (72%). The two control group schools that dropped out during the trial were located in regions 1 and 2 which will account for some of this regional variation. In all, response rates varied between 62% (Y7, region 1 control) and 86% (Y8, region 1 intervention).

In summary, response was pretty high, generally higher for the intervention compared with the control group and that it varied by region and year group. What remains unknown is whether and how this non-response (or attrition) had an impact on the balance of the (control and intervention group) samples. In other words, to assess whether this 25% non-response introduced any statistical bias (or imbalance) into the samples.

7.2.2 Sample Balance & Attrition

To examine how balanced (or similar) the MRP intervention and control group samples were, baseline statistics are compared at two points in time; baseline (October 2013) and final (September 2014). This comparison was done at the pupil and school levels. Specifically, the MRP intervention and control group samples are compared in terms of; pupil attainment in KS2 maths, pupil FSM, pupil gender, pupil FSM, school GCSE attainment, school admissions policy, school OFSED ratings, school percentage of pupils classed as FSM and school IDACI score.

Sample Balance & Attrition - KS2 Maths Attainment

Table 14 summarises the mean KS2 maths attainment for the control and intervention group samples. This is done for the baseline sample and the final sample. Alongside the mean KS2 maths attainment statistics, hedges g effect size statistics are shown.

At baseline, a notable pupil-level attainment imbalance is observed - on average the control group sample attained significantly higher in KS2 maths compared with the intervention group sample (an effect size of g= - 0.25 overall). This pattern is replicated across year groups but is strongest within Y8 (g= -0.39) and weakest in Y7 (g= -0.16). Looking across the regions, it is apparent that the imbalance relates primarily to regions 1 and 2 and not to region 3. The strongest imbalance is seen in Y8 for region 2 (g=-0.69).

These patterns are also observed in the final sample and attrition does seem to have made them more pronounced (the effect size increased to g=-0.29 overall). The imbalance remains present across year groups but strongest in Y8 (g=-0.46) and weakest in Y7 (g=-0.17). The imbalance remains concentrated in regions 1 and 2 with no statistical imbalance seen in region 3. The strongest imbalance is again seen in Y8 for region 2 (g=-0.87).

Finding an imbalance between the control and intervention group samples will mean that it is important that the eventual statistical models take this into account. This will be done by including KS2 maths attainment as a pupil-level explanatory variable. This way, the models will be focusing on progress between KS2 maths and the PiM outcome measures.

To provide a visual overview of the pre-test KS2 maths fine points score measure, Figures 4 to 7 provide a series of histograms. KS2 maths score histograms for intervention and control group samples are shown for the baseline sample alongside the final sample. Figure 4 shows this for Y7-Y9 samples combined and Figures 5 to 7 show this for the Y7, Y8 and Y9 samples separately.

From this KS2 maths histograms, floor and ceiling effects are evident - most strikingly in Y7. A suggestion of a negative skew is also evident. Neither of these brings large concerns for the use of KS2 maths attainment as an explanatory variable. Looking more closely at Y7 and Y8 - a cluster of cases are found at the '6.5' maximum grade - whilst no such cluster exists for the Y9 sample.

The Y7 sample will have taken KS2 maths in the summer of 2013 - 275 (9%) have a KS2 fine point maximum score of 6.5. The next highest score is 6.0 which leaves a gap within the KS2 Maths attainment distribution. The Y8 sample will have taken KS2 maths in summer 2012 - 119 (4.5%) have the 6.5 maximum and the gap between this and 6.0 is also evident. The Y9 sample will have taken KS2 maths in summer 2011 and the maximum value of this (continuous) distribution is 6.0. It seems that the 6.5 level was introduced in 2012.

Table 14 Mean KS2 Maths Attainment

Comparing control and intervention group samples

Baseline and Final

	At baseline October 2013 n=8,430[1]			Final September 2014 n=6,565		
	Control	**Intervention**	**Effect Size (hedges g)**	**Control**	**Intervention**	**Effect Size (hedges g)**
All Areas Combined						
Y7 to Y9 combined	5.0	4.8	- 0.25*	5.0	4.8	- 0.29*
Y7	5.0	4.9	- 0.16*	5.1	4.9	- 0.17*
Y8	5.0	4.6	- 0.39*	5.0	4.7	- 0.46*
Y9	4.9	4.7	- 0.24*	5.0	4.8	- 0.29*
Region 1						
Y7 to Y9 combined	5.0	4.8	- 0.17*	5.1	4.9	- 0.27*
Y7	4.9	4.9	+ 0.03	5.0	5.0	0.00
Y8	5.0	4.7	- 0.45*	5.1	4.7	- 0.58*
Y9	5.0	4.8	- 0.23*	5.1	4.8	- 0.38*
Region 2						
Y7 to Y9 combined	5.1	4.6	- 0.54*	5.2	4.7	- 0.62*
Y7	5.2	4.8	- 0.49*	5.3	4.9	- 0.52*
Y8	5.0	4.4	- 0.69*	5.2	4.4	- 0.87*
Y9	5.0	4.6	- 0.52*	5.1	4.7	- 0.58*
Region 3						
Y7 to Y9 combined	4.9	4.8	- 0.07	4.9	4.9	- 0.01
Y7	4.9	4.8	- 0.16*	5.0	4.9	- 0.08
Y8	4.9	4.9	- 0.01	4.9	5.0	+ 0.06
Y9	4.8	4.8	0.00	4.8	4.8	+ 0.04

** - statistically significant (p<0.05).*

[1] *- Whilst at baseline there were a total of 8,777 potential participants, KS2 maths attainment was only available for 8,430 (96%) of these.*

What these analyses affirm is that an outcome only analysis would not be suitable for this RCT. KS2 maths attainment needs to be included in the model at the pupil level to ensure that the initial observed imbalance is controlled for in the multilevel analysis; this shifts the analysis focus from one of raw attainment and towards one of progress between two attainment points (i.e. KS2 and PiM).

Figure 4 KS2 Maths Attainment for combined Y7 to Y9 sample (Baseline & Final)

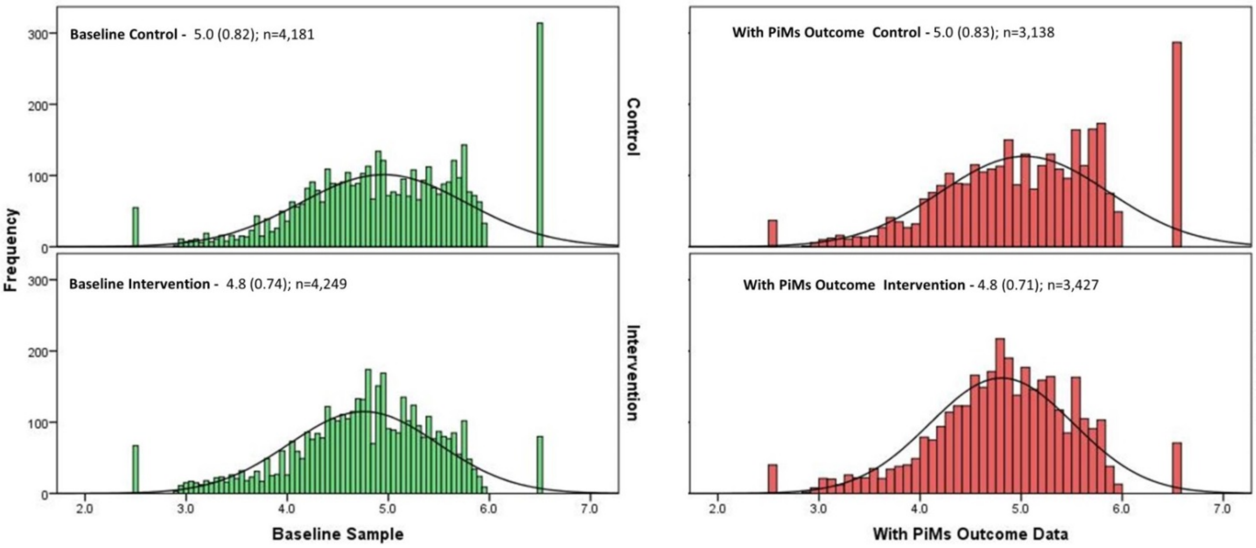


Figure 5 KS2 Maths Attainment for Y7 subsample (Baseline & Final)

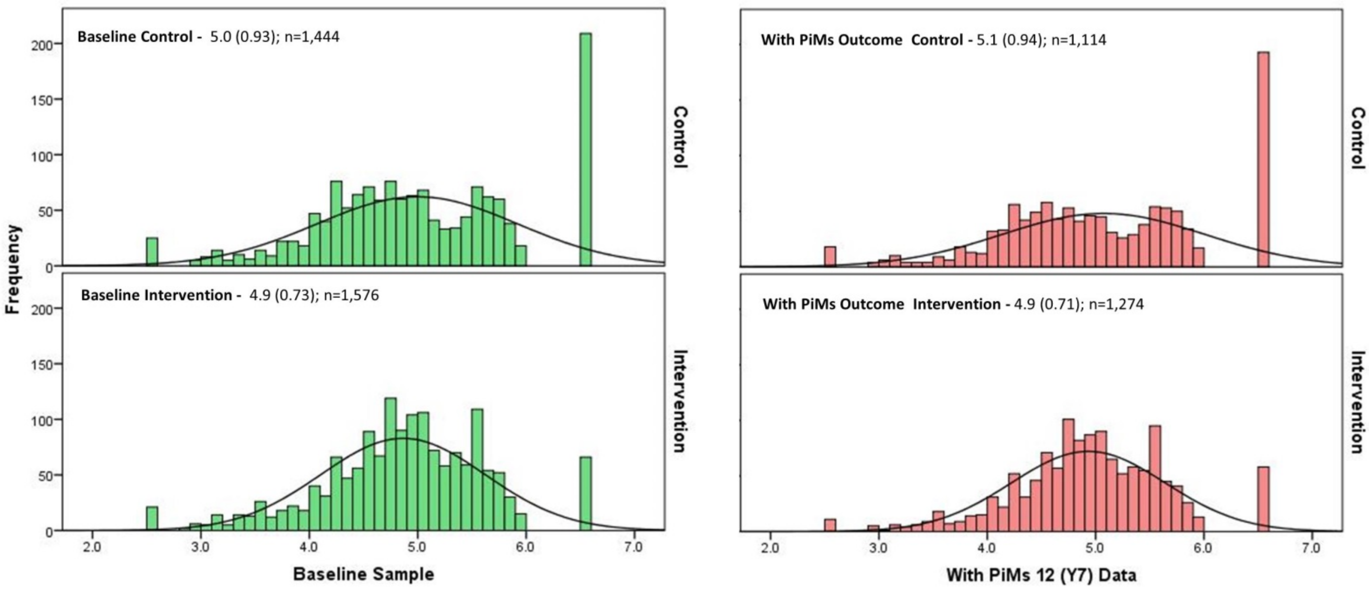


Figure 6 KS2 Maths Attainment for Y8 subsample (Baseline & Final)

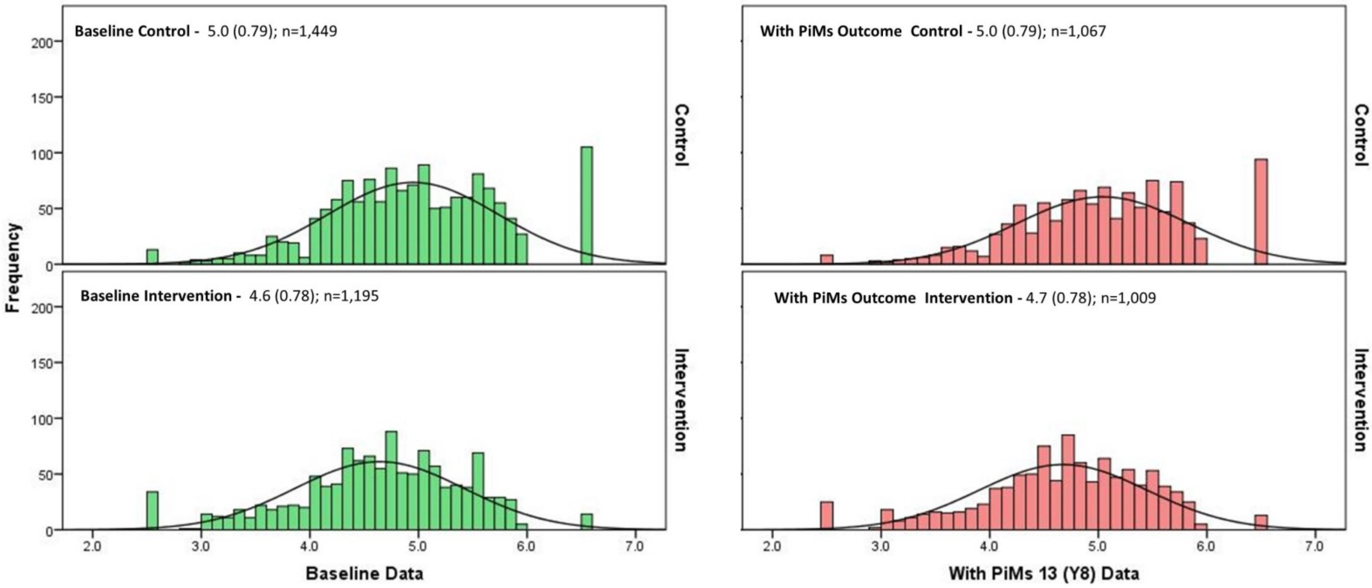
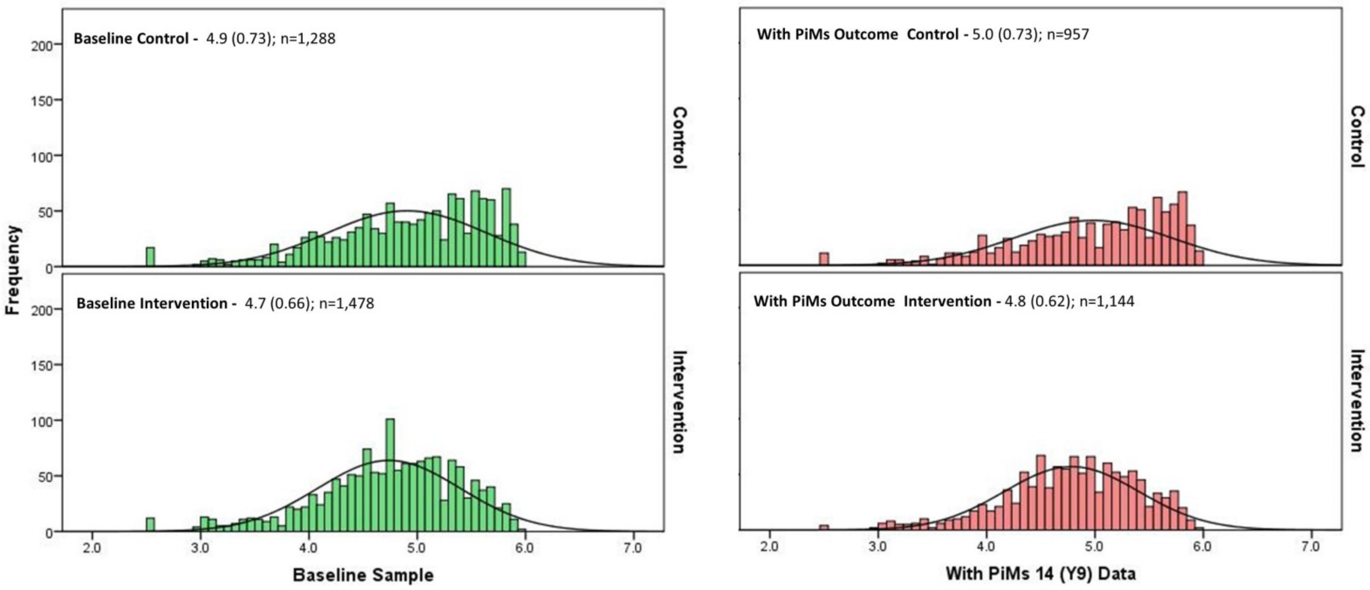


Figure 7 KS2 Maths Attainment for Y9 subsample (Baseline & Final)



Sample Balance & Attrition - Pupil Demographics (FSM, gender & SEN)

The intervention and control group samples were profiled at baseline in terms of their gender balance, proportion of FSM participants and the proportion with a SEN. These analyses are replicated with the final sample and the sample. Tables 15 to 16 provide a summary of these analyses.

As shown in Table 15, in terms of **FSM**, at baseline the intervention group sample were slightly more likely to be FSM (27%) compared with control group sample

(24%). This difference is strongest in Y9 and particularly Y9 within region 1 and 2. The patterns observed within the final sample largely reflect those found at the previous time points. In summary, the FSM imbalance observed is small and by including FSM as a pupil-level explanatory variable in the model, this will be accounted for in the analyses.

Table 15 Percentage classified as 'FSM'

Comparing control and intervention group samples

Baseline and Final

	At baseline October 2013 n=8,431[1]			Final September 2014 n=6,565		
	Control	**Intervention**	**Effect Size (hedges g)**	**Control**	**Intervention**	**Effect Size (hedges g)**
All Areas Combined						
Y7 to Y9 combined	24%	27%	+ 0.06*	22%	25%	+ 0.06*
Y7	25%	25%	0.00	23%	23%	- 0.01
Y8	26%	28%	+ 0.05	23%	26%	+ 0.08
Y9	22%	28%	+ 0.14*	20%	26%	+ 0.14*
Region 1						
Y7 to Y9 combined	19%	21%	+ 0.06	17%	20%	+ 0.10*
Y7	22%	19%	- 0.07	18%	18%	0.00
Y8	19%	22%	+ 0.07	17%	21%	+ 0.10
Y9	15%	24%	+ 0.23*	14%	22%	+ 0.21*
Region 2						
Y7 to Y9 combined	28%	31%	+ 0.07	24%	29%	+ 0.10*
Y7	30%	30%	0.00	29%	27%	- 0.04
Y8	32%	34%	+ 0.04	25%	33%	+ 0.17*
Y9	23%	31%	+ 0.18*	20%	28%	+ 0.20*
Region 3						
Y7 to Y9 combined	27%	28%	+ 0.01	25%	25%	0.00
Y7	25%	29%	+ 0.09	24%	25%	+ 0.01
Y8	28%	27%	- 0.02	26%	24%	- 0.03
Y9	28%	27%	- 0.02	25%	26%	+ 0.02

** - statistically significant (p<0.05).*

[1] - Whilst at baseline there were a total of 8,777 potential participants, FSM status details was only available for 8,431 (96%) of these.

As shown in Table 16, in terms of **gender**, at baseline a slight male bias in both the control and intervention group samples is observed overall and in most regions and year groups. This male bias is slightly stronger in the intervention group (45% female overall) compared with the control group (48% female overall). There is some fluctuation in the gender balance of control and intervention group samples over the regions / year groups - the largest difference is seen in Y9 and particularly in region 3 where the proportion of females is 57% and 46% for the control and intervention group samples respectively. Looking at the final sample and the impact of non-response on the gender balance of the two samples, things largely remain similar to that seen at baseline. The inclusion of a pupil-level gender explanatory variable in the models will be needed to take account of the observed (slight) gender imbalance observed at baseline.

Table 16 Gender: Percentage female - Comparing control and intervention group samples Baseline and Final

	At baseline October 2013 n=8,473[1]			Final September 2014 n=6,565		
	Control	**Intervention**	**Effect Size (hedges g)**	**Control**	**Intervention**	**Effect Size (hedges g)**
All Areas Combined						
Y7 to Y9 combined	48%	45%	- 0.06*	49%	46%	- 0.06*
Y7	44%	45%	+ 0.02	44%	44%	0.00
Y8	49%	45%	- 0.08	51%	46%	- 0.10*
Y9	52%	46%	- 0.12*	52%	47%	- 0.10*
Region 1						
Y7 to Y9 combined	47%	41%	- 0.12*	48%	40%	- 0.16*
Y7	47%	41%	- 0.12	46%	38%	- 0.16*
Y8	49%	39%	- 0.20*	51%	40%	- 0.22*
Y9	47%	42%	- 0.10	47%	42%	- 0.10
Region 2						
Y7 to Y9 combined	46%	48%	+ 0.04	47%	48%	+ 0.02
Y7	40%	47%	+ 0.14*	42%	48%	+ 0.12
Y8	46%	47%	+ 0.02	46%	46%	0.00
Y9	53%	49%	- 0.08	53%	50%	- 0.06
Region 3						
Y7 to Y9 combined	51%	48%	- 0.06	50%	50%	0.00
Y7	45%	48%	+ 0.06	43%	48%	+ 0.10
Y8	51%	51%	0.00	53%	51%	- 0.04
Y9	57%	46%	- 0.22*	54%	49%	- 0.10

** - statistically significant (p<0.05).*

[1] - Whilst at baseline there were a total of 8,777 potential participants, gender details were only available for 8,473 (97%) of these.

As shown in Table 17, in terms of In terms of **SEN**, at baseline the intervention group sample are seen to be more likely to have a SEN (21%) compared with the control group sample (17%). This imbalance is located primarily within the Y8 and Y9 samples and within region 1 and 2. The largest difference is seen in Y8 for region 1 (14% of the control sample compared with 26% of the intervention sample having a SEN) and for region 2 (21% of the control sample compared with 31% of the intervention sample having a SEN). Looking at the final sample, the SEN imbalances observed at baseline remain largely similar. The inclusion of a pupil-level SEN explanatory variable in the models will be needed to take account of the observed (slight) SEN imbalance observed at baseline.

Table 17 Special Educational Needs: Percentage with SEN - Comparing control and intervention group samples

Baseline and Final

	At baseline October 2013 n=8,431[1]			Final September 2014 n=6,565		
	Control	**Intervention**	**Effect Size (hedges g)**	**Control**	**Intervention**	**Effect Size (hedges g)**
All Areas Combined						
Y7 to Y9 combined	17%	21%	+ 0.10*	15%	18%	+ 0.08*
Y7	18%	19%	+ 0.03	16%	16%	- 0.02
Y8	18%	25%	+ 0.17*	16%	23%	+ 0.19*
Y9	15%	19%	+ 0.12*	13%	17%	+ 0.10*
Region 1						
Y7 to Y9 combined	15%	21%	+ 0.15*	15%	21%	+ 0.15*
Y7	17%	19%	+ 0.06	17%	17%	+ 0.02
Y8	14%	26%	+ 0.24	15%	29%	+ 0.33
Y9	13%	18%	+ 0.16	12%	17%	+ 0.14
Region 2						
Y7 to Y9 combined	17%	23%	+ 0.14*	15%	21%	+ 0.15*
Y7	17%	19%	+ 0.05	17%	17%	+ 0.02
Y8	21%	31%	+ 0.24*	15%	29%	+ 0.33
Y9	15%	21%	+ 0.16	12%	17%	+ 0.14
Region 3						
Y7 to Y9 combined	19%	18%	- 0.02	17%	15%	- 0.06
Y7	20%	20%	- 0.01	16%	14%	- 0.06
Y8	20%	16%	- 0.10	18%	14%	- 0.12
Y9	17%	18%	+ 0.04	16%	17%	+ 0.01

** - statistically significant (p<0.05).*

[1] - Whilst at baseline there were a total of 8,777 potential participants, SEN details were only available for 8,431 (96%) of these.

The observed imbalances between the control and intervention group samples with respect to FSM, gender and SEN are all relatively small when compared to the KS2 attainment imbalance. By including pupil-level explanatory variables into the model, these imbalances will be statistically taken into account within the impact analyses.

Whilst the imbalance can statistically be accounted for, it is worth considering the assumptions behind this. The key imbalance is one of KS2 maths attainment with the control group attaining higher on average compared with the intervention group. By including the KS2 explanatory variable into the model this difference will be accounted for and the focus will shift from being on 'attainment in PiM' to one of progress between KS2 and PiM. This does assume that the rate of progress does not depend upon initial attainment. In other words, the assumption is that under random conditions, progress of pupils with relatively high KS2 maths attainment would be similar (parallel) to the progress of pupils with relatively low KS2 maths attainment. This assumption allows the influence of 'group' (control or intervention) to be isolated as a 'causal' influence on this rate of progress. Another point worth noting is that KS2 maths attainment will be more recent for the Y7 cohort (around 6 months prior to randomisation) compared with the Y9 cohort (around 2 and half years prior to randomisation).

Sample Balance & Attrition - School Context

In terms of **school-level attainment**, 85% of the control group and 82% of the intervention group were at schools where over 59% of pupils attained 5+ A*-C GCSEs. There is some fluctuation in this pattern across year groups and regions. The largest difference is seen in Y8 of region 3 where 96% of the control group and 82% of the intervention group were at schools with this level of attainment. Conversely, in Y7 of region 1, 65% of the control group and 80% of the intervention group were at schools with this level of attainment. A school level GCSE attainment variable will be included into the models to address these imbalances.

In terms of **selective admissions policies**, 13% of the control group and 0% (none) of the intervention group were at schools that used academic selection as part of their admissions policies. This key school level imbalance will be addressed by including a school-level dummy variable that identifies whether a school is selective (=1) or not. This dummy will assess whether progress of pupils within these three selective control schools is significantly different to other control group schools. If this is the case, the model will isolate this difference as a significant coefficient in the model.

In terms of **recent OFSTED ratings**, 77% of the control group and 78% of the intervention group were at schools that had an OFSTED rating of 1 or 2. There is some fluctuation in across year groups and regions. The largest difference is seen in Y7 of region 1 where 61% of the control group and 88% of the intervention group were at schools with this OFSTED rating. Conversely, in Y9 of region 2, 85% of the control group and 68% of the intervention group were at schools with this rating. A school level OFSTED variable will be included into the models to address these slight imbalances.

In terms of **school level %FSM**, the weighted average of %FSM amongst control group participants was 13% compared with 14% amongst the intervention group. There is some fluctuation in across year groups and regions. The largest difference is seen in Y9 of region 1 where the weighted averages were 8% for the control group and 14% for the intervention group. Conversely, in Y8 of region 3, the averages were 18% for the control group and 15% for the intervention group. A school level %FSM variable will be included into the models to address these slight imbalances.

In terms of **school level IDACI**, the weighted average of IDACI amongst control group participants was 0.14 compared with 0.16 amongst the intervention group. There is some fluctuation in across year groups and regions. The largest difference is seen in Y9 of region 1 where the weighted averages were 0.08 for the control group and 0.22 for the intervention group. Conversely, in Y8 of region 3, the averages were 0.21 for the control group and 0.17 for the intervention group. A school level IDACI variable will be included into the models to address these slight imbalances.

These school level analyses have helped to identify key school-context variables that will be included into the models. The most striking findings are around school admissions and GCSE attainment.

7.2.3 Teacher sample

Of the originally nominated 120 core teachers:

104 (87%) had an initial teaching qualification in mathematics.

Table 18 provides detail of data of qualification

Table 18 Qualification Data

Year qualified	% of teachers (total n = 119)
Before 1999	21
2000-2005	13
2006-2009	19
2010	12
2011	10
2012	20
2013	6

Table 19 provides detail of the roles of teachers.

Table 19 Roles of Teachers

Role in Department	% of teachers (total n=115)
Head/Deputy Head of Department/Mathematics/KS3	5
Leadership role (subject leader, KS3 Co-ordinator)	37
Mathematics teacher	47
NQT	7
Second in Mathematics	4

7.2.4 The PiM Outcome Variables

Table 20 summarises the mean PiM scores for intervention and control group samples and Figure 8 presents the distributions of the PiM standardised age scores for the combined Y7 to Y9 sample and within each of the three year groups.

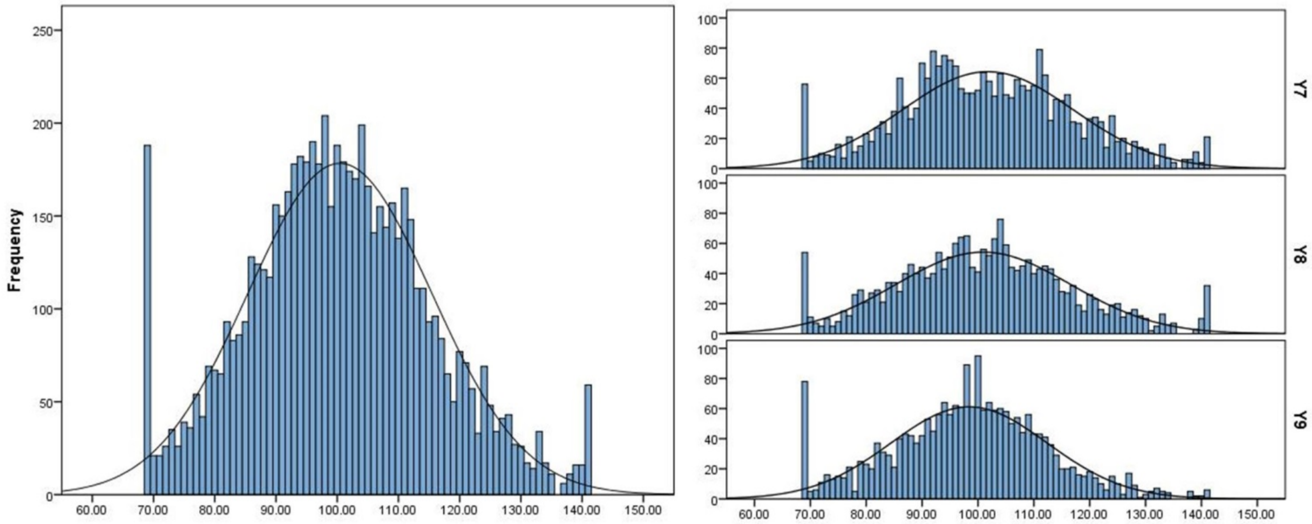
Table 20 Mean PiM Standardised Age Scores

Year Group (Outcome)	Control Group	Intervention Group	Mean Difference	Effect Size (Hedges g)
Y7 (PiM12)	103.1	100.1	-3.0	-0.20*
Y8 (PiM13)	104.0	97.6	-6.4	-0.43*
Y9 (PiM13)	101.6	95.7	-5.9	-0.43*
Y7to9 (PiM12to14)	103.0	97.9	-5.1	-0.35*

** statistically significant p<0.05*

On average, the control group sample attained significantly higher on the PiM tests compared with the intervention group. However, given that an attainment bias towards the control group was also observed with the KS2 maths prior attainment measure (see Table 14), what might be seen in Table 20 is an echo of this prior-attainment imbalance. In order to assess whether the difference increased (or widened) between the baseline (KS2 maths) and the outcome (PiM score) requires a statistical model that accounts for the initial imbalance and focused on capturing progress between these two attainment points.

Figure 8 Histograms for PiM Standardised Age Scores Y7 to Y9 combined and within each year group



7.2.5 The relationship between KS2 Maths and the PiM Outcome Variables

Table 21 summarises the Pearson correlation statistics between KS2 maths prior attainment and the PiM outcome measures for intervention and control group samples.

Table 21 Mean PiM Standardised Age Scores

Year Group (Outcome)	Control Group	Intervention Group
Y7 (PiM12)	0.87*	0.79*
Y8 (PiM13)	0.84*	0.80*
Y9 (PiM13)	0.81*	0.71*
Y7to9 (PiM12to14)	0.84*	0.77*

** statistically significant p<0.05*

As was anticipated, a strong positive correlation between KS2 maths and the PiM outcome measures is observed. The strength of this correlation diminishes with age with Y7 and Y8 observed to have a stronger correlation compared with Y9.

7.3. Intent to treat analysis - mathematics attainment

Table 22, below, summarises the intention to treat multilevel analyses for the four PiM outcome variables. More details on the models behind the effect sizes shown in Table 22 can be found in Appendix A1. Additionally, Appendix A provides some guidance on how to read the model tables and details on how the effect size statistics shown in Table 22 were calculated. Overall, and within each of the KS3 year groups, a negative effect size is observed. This indicates that, on average and once KS2 maths attainment and other statistically significant explanatory variables have been taken into account, the intervention group attain a lower score on the PiM test compared with the control group. However, in all cases, the negative effect size is not statistically significant. This means that whilst we have observed negative effect sizes within our sample, we are unable to reliably distinguish them from 'zero'.

Therefore, from these 'intention to treat' analyses, we conclude that the MRP PD programme had no statistically significant impact on pupil attainment in Y7, Y8 or Y9 or for the combined Y7 to Y9 sample.

Table 22 PiM Outcome Measures - All regions combined (Intention to treat)

Main Effects; Intervention v Control, Hedges g effect sizes & 95% CIs

Please see Appendix E2 for full details on these four multilevel models.

Year Group (Outcome)	No of Schools	No of Classes	No of Pupils	Effect size (Hedges g)	95% Confidence Intervals (CI)	Comment
Y7 (PiM12)	55	128	2,388	-0.02	-0.12 to +0.08	Not Significant
Y8 (PiM13)	56	110	2,076	-0.08	-0.24 to +0.07	Not Significant
Y9 (PiM13)	52	111	2,101	-0.11	-0.29 to +0.06	Not Significant
Y7to9 (PiM12to14)	58	349	6,565	-0.07	-0.15 to +0.02	Not Significant

7.4 MR items only

Tables 23 to 24 below summarise the intention to treat analyses for the three PiM subscales. The first is based on items identified as being related to multiplicative reasoning in but not necessarily linked to the topics covered in the MRP (subscale 1); the second is based on items identified as being related to multiplicative reasoning and also identified as being weakly or strongly related to topics covered within the MRP intervention (subscale 2) and the final subscale is based on items identified by Andrew Izsák as being related to multiplicative reasoning and identified as being strongly related to topics covered within the MRP intervention (subscale 3). These subscales are shown only within the three year groups (and only in Y8 and Y9 for subscale 3 because the PiM 12 test for Y7s contained no items deemed to be 'strongly' connected to the MRP).

In all cases, the effect sizes are not statistically significant and so the overall conclusion for the intention to treat analyses is one of 'no impact'.

Specifically, no evidence was found to show that the MRP teacher PD intervention had a positive or negative impact on pupil attainment in maths generally and pupil attainment relating to test items related to multiplicative reasoning during the six month trial period.

Table 23 PiM Outcome Subscale 1 - (MR related)

All regions combined (Intention to treat)

Main Effects; Intervention v Control, Hedges g effect sizes & 95% CIs

Please see Appendix E3 for full details on these multilevel models.

Year Group / Outcome	No of Schools	Number of classes	Number of Pupils	Effect size (Hedges g)	95% Confidence Intervals (CI)	Comment
Y7 (PiM12)	55	128	2,388	0.00	-0.11 to +0.12	Not Significant
Y8 (PiM13)	56	2,076	2,076	-0.02	-0.21 to +0.17	Not Significant
Y9 (PiM13)	52	2,101	2,101	-0.15	-0.33 to +0.03	Not Significant

Table 24 PiM Outcome Subscale 2 - (MR related & weakly / strongly connected)

All regions combined (Intention to treat)

Main Effects; Intervention v Control, Hedges g effect sizes & 95% CIs

Please see Appendix E3 for full details on these multilevel models.

Year Group / Outcome	No of Schools	Number of Pupils	Effect size (Hedges g)	95% Confidence Intervals (CI)	Comment
Y7 (PiM12)	55	2,388	0.00	-0.09 to +0.09	Not Significant
Y8 (PiM13)	56	2,076	+0.01	-0.17 to +0.18	Not Significant
Y9 (PiM13)	52	2,101	-0.09	-0.23 to +0.05	Not Significant

Table 25 PiM Outcome Subscale 3 - (MR related & strongly connected)

All regions combined (Intention to treat)

Main Effects; Intervention v Control, Hedges g effect sizes & 95% CIs

Please see Appendix E3 for full details on these multilevel models.

Year Group / Outcome	No of Schools	Number of Pupils	Effect size (Hedges g)	95% Confidence Intervals (CI)	Comment
Y8 (PiM13)	56	2,076	+0.02	-0.14 to +0.18	Not Significant
Y9 (PiM13)	52	2,101	-0.08	-0.19 to +0.04	Not Significant

7.5 Fidelity analysis

Trial fidelity relates to whether the 'intervention' was experienced across participating schools and classrooms and by participating pupils in a way that was consistent (or similar) and in a way that was anticipated / intended by the delivery partners.

For the 'intention to treat' analyses summarised in sections 7.3 and 7.4 above, trial fidelity is ignored. If an intervention is found to have a statistically significant impact within an 'intention to treat' approach, it provides the best strength of evidence that a similar impact would be seen if the intervention were rolled out more widely. This is because 'in the wild' it is likely that there would be similar variations in fidelity - in how the intervention was received by teachers and pupils. If an intervention can withstand such variations in a trial within 'intention to treat' conditions, the evidence that this will manifest more widely is maximised.

However, ignoring fidelity results in a partial picture. This was a pilot trial of a teacher-focused CPD intervention that aimed to result in measurable impact in pupil attainment in a nine month period. An examination of how pupils and teachers experienced the trial is hence valuable to help build a more complete picture of the MRP intervention. These analyses are summarised here.

The approach we adopted for capturing fidelity draws on teacher engagement with the MRP intervention and practical details on how the trial was experienced by pupils.

Fidelity in terms of teacher engagement relates only to the intervention group sample and how engaged their core mathematics teachers were with the MRP intervention. Fidelity in terms of practical details relates to how both the intervention and control group samples experienced the trial.

7.5.1 Teacher engagement

Two aspects of teacher engagement are drawn on:

- attendance at TIME events
- reported use of the MRP unit materials

In terms of attending TIME event days, 60% of intervention group pupils had teachers who attended all four days, 25% had teachers who attended for three days, 10% for two days and 5% for one day (see section 8.2.1 for further details of attendance).

As shown in Table 26, attendance of core teachers at TIME events were found to be positively correlated with the PiM outcome measures (overall and the three subscales). On average, pupils who were taught by a teacher who attended all four TIME days attained higher in PiM (97.8) compared with pupils taught by a teacher who only attended just one of the TIME days (92.6) - which equates to an effect size difference of about (g=) +0.39 standard deviations.

Table 26 Attendance of TIME events and Mean PiM attainment

Intervention Schools Only

3,427 pupils in 174 classes in 30 schools

	1 day	2 days	3 days	4 days	Effect size (hedges g) 4 days v 1 day
Overall PiM Attainment	92.6	98.1	98.9	97.8	+0.39
PiM Subscale 1 (MR related)	9.8	13.5	14.5	13.3	+0.47
PiM Subscale 2 (MR related & weakly or strongly connected to the MRP)	3.5	4.0	5.0	4.2	+0.20
PiM Subscale 3 (MR related & strongly connected to the MRP)	1.3	1.9	2.4	1.9	+0.32

In terms of use of the MRP materials, 22% of intervention group pupils were taught by teachers who did not report to use any of the MRP materials, 43% had teachers who reported to use materials from 1 to 3 units, 26% had teachers who reported to use materials from 4 to 6 units and 9% had teachers who reported to use materials from 7 or more units.

As shown in Table 27, use of the MRP materials was found to be positively correlated with the PiM outcomes. On average, the 78% of pupils who were taught by a teacher who reported to use some of the MRP materials attained higher in PiM (98.5) compared with the 22% of pupils taught by a teacher who did not report to use the materials (95.7) - which equates to an effect size difference of about (g=) +0.2.

On average, the 9% of pupils taught by a teacher who reported to have used material from at least 7 out of the 9 possible units attained higher still in PiM (103.0 - an effect size of g=+0.6 compared with 'no reported use').

Table 27 Use of MRP materials and Mean PiM attainment

Intervention Schools Only

3,427 pupils in 174 classes in 30 schools

	No Reported use	1-3 units	4-6 units	7+ units	Effect size (hedges g) 7+ units v no reported use
Overall PiM Attainment	95.7	98.5	97.0	103.0	+0.60
PiM Subscale 1 (MR related)	12.0	14.0	12.9	16.2	+0.61
PiM Subscale 2 (MR related & weakly or strongly connected to the MRP) PiM12to13 Combined (Y7 to Y9)	3.5	4.8	4.1	4.9	+0.50
PiM Subscale 3 (MR related & strongly connected to the MRP) PiM12 to13 Combined (Y7 to Y9)	1.4	2.4	1.9	2.2	+0.42

From the empirical evidence, it seems that teacher engagement with the MRP intervention in terms of attending TIME events and reported use of the MRP materials are both statistically associated with higher levels of attainment on the PiM outcome measures (i.e. both are positively correlated).

Drawing on both TIME event attendance and use of the MRP materials, a criterion was created to directly distinguish teachers that attended and used materials from teachers who did not. In terms of attendance, teachers who attended at least 2 of the 4 TIME event days were selected. In terms of the use of the MRP materials, teachers who reported to use at least some of the MRP materials (1+ units) were selected.

Combining these two teacher-engagement criteria results in identifying a subsample of 805 (23%) pupils in the intervention group schools. These 805 pupils were taught by teachers with low / no engagement with the MRP in terms of attendance and material use. The remaining 2,622 (77%) of pupils in the intervention group schools all were taught by a teacher who attended at least half of the TIME event days and reported to use some of the MRP materials.

In terms of PiM attainment, on average this 77% 'teacher engaged' intervention group subsample attained higher (98.6) compared with the remaining 23% 'teacher not engaged' intervention group subsample (95.5) - a difference with an effect size of g=+0.2.

7.5.2 Practical Fidelity

Two aspects of practical details are drawn on:

- Whether a pupil moved classes during the trial period
- Whether a pupil was taught solely by one of the named core maths teachers throughout the trial.

Practical fidelity relates to both the intervention and control group samples. In terms of pupil movement, 4% of the control group sample and 6% of the intervention group sample moved during the trial period. Table 28 summarises the mean PiM attainment for intervention and control group samples of pupils who moved class compared with pupils who did not move.

Table 28 Pupil Movement and Mean PiM attainment

Intervention and Control Schools

6,656 pupils in 349 classes in 58 schools

	Did not move	Moved class	Effect size (hedges g) Move - Not move
Overall PiM Attainment			
Intervention school sample	97.9	98.0	+ 0.01
Control schools	103.2	97.9	- 0.33
PiM Subscale 1 (MR related)			
Intervention schools	13.5	12.7	- 0.11
Control schools	16.6	13.4	- 0.34
PiM Subscale 2 (MR related & weakly or strongly connected to the MRP)			
Intervention schools	4.4	3.7	- 0.20
Control schools	5.6	3.5	- 0.50
PiM Subscale 3 (MR related & strongly connected to the MRP)			
Intervention schools	2.0	1.6	- 0.22
Control schools	2.8	1.8	- 0.42

Table 28 shows that within the intervention group sample, the average overall PiM attainment for the 6% of pupils who moved (98.0) was very similar to the 94% who did not move (97.9). This difference was wider within the three PiM subscales. Pupil movement is also seen to be associated with lower levels of attainment in PiM amongst the control group subsample (to a greater extent to that seen with the intervention group sample).

In summary, whilst we have no details on why pupils move class[33], within both the intervention and control group samples, pupil movement is associated with lower levels of attainment across the PiM measures.

In addition to pupils moving away from their original mathematics class during the period of the trial, the fidelity analyses considered core teachers themselves. Ninety-three percent of the control group sample and 99% of the intervention group sample were taught maths solely by a core maths teacher who was named prior to randomisation in October 2013[34]. The rest were taught by multiple teachers (one of which was usually the core). Table 29 summarises the mean PiM attainment for intervention and control group samples of pupils taught solely by a named core maths teacher compared with pupils taught by multiple teachers.

As shown in Table 29 within the intervention group sample, the average PiM attainment for the 99% of pupils taught solely by a core teacher (98.1) was considerably higher compared with the 1% of pupils taught by more than one teacher (80.7) - an effect size of -1.3. Within the control group sample, whilst a greater percentage was taught by multiple teachers, the impact on PiM is less evident than what is observed with the intervention group sample.

[33] For example, pupils might move between ability groups or leave a school etc.

[34] The trial did actually experience a greater amount 'teacher movement' but in nearly all other cases where this occurred, the school decided not to test the pupils. Of those pupils with complete data, only 1% of the intervention group and 7% of the control group were not taught solely by the named core maths teacher.

Table 29 Core Teachers and Mean PiM attainment

Intervention and Control Schools

6,656 pupils in 349 classes in 58 schools

	Taught solely by named core maths teacher throughout the trial period.	Taught by multiple teachers in maths.	Effect size (hedges g) Multiple teachers - Sole core teacher
Overall PiM Attainment			
Intervention school sample	98.1	80.7	- 1.34
Control schools	103.1	100.7	- 0.15
PiM Subscale 1 (MR related)			
Intervention schools	13.6	4.5	- 1.21
Control schools	16.5	15.9	- 0.07
PiM Subscale 2 (MR related & weakly or strongly connected to the MRP)			
Intervention schools	4.4	0.9	- 1.02
Control schools	5.5	6.1	+ 0.14
PiM Subscale 3 (MR related & strongly connected to the MRP)			
Intervention schools	2.1	0.4	- 0.76
Control schools	2.7	3.8	+ 0.42

These two practical details were brought together to create a practical fidelity criteria. This criteria identified an intervention school subsample of 93% (n=3,187) and a control group subsample of 89% (n=2,791) of participating pupils who did not move classes during the trial and were taught solely by the named core maths teacher throughout the trial.

On average, the intervention and control group subsamples who did not move classes during the trial and were taught solely by the named core maths teacher throughout the trial attained slightly higher on the PiM outcome - an effect size of g=+0.2 for both.

7.5.3 Bringing teacher engagement and practical fidelity together

The final stage is to draw the teacher engagement and practical aspects of fidelity together. This identifies an intervention group subsample of 72% (n=2,462) who did not move classes during the trial, were taught solely by the named core maths teacher who attended at least half of the TIME event days and reported to use at least some of the MRP material. On average, this 72% subsample attained higher on the PiM outcome (99.0) compared with the remaining 28% (95.1) - an effect size of g=+0.3.

The practical details drawn on for the intervention group sample were also applied to the control group sample. This is to try to ensure a like by like comparison. Specifically, this identifies a control group subsample of 89% (n=2,791) of pupils who did not move during the period of the trial and were taught solely by the named core maths teacher. On average, this 89% control group subsample attained higher on the PiM outcome (103.4) compared with the remaining 11% (99.6) - an effect size of g=+0.2.

These intervention and control group subsamples will be the focus for analyses presented in section 7.6 below - the 'on treatment' analyses.

7.6 On treatment analysis

The 'on treatment' analyses draws on the fidelity analyses summarised in section 7.5 to focus the statistical 'impact' analyses on a subsample of pupils in both the intervention and control group schools.

The main (intention to treat) analyses summarised in sections 7.3 and 7.4 above, involved data for 6,565 pupil participants taught in 349 Y7, Y8 and Y9 classes in 58 secondary schools and ignores issues of trial fidelity. The 'on treatment' analysis introduced minimum standard criteria regarding trial fidelity in terms of pupil movement and teacher engagement that resulted in reducing the sample to 5,253 pupil participants in 269 Y7, Y8 and Y9 classes in 56 secondary schools.

Specifically, the samples were adapted using the following criteria:

Practical fidelity: control and intervention group pupils remained in same class they were in at the point of randomisation and were taught solely by the named core maths teacher throughout the trial.

Teacher Engagement: intervention group pupils were taught by a core maths teacher who attended at least 2 of the 4 TIME event days and reported to use some of the MRP materials.

Table 30 Intention to treat and on treatment

	Intervention sample	Control sample
Intention to treat analyses (sections 7.3 & 7.4)	3,427	3,138
On treatment analyses (practical fidelity AND teacher engagement)	2,462 (72%)	2,791 (89%)

Table 31 summarises the 'on-treatment' analyses for the four main PiM outcome measures and Tables 32, 33 and 34 provide similar 'on-treatment' analyses summaries for the PiM subscale measures.

These analyses re-affirm the conclusion that the MRP teacher PD intervention did not result in a statistically significant impact on pupil attainment (as measured by the series of PiM outcomes) during the six month trial period.

Table 31 PiM Outcome Measures - Overall PiM attainment

1.1.1.1.2.1 All regions combined (On Treatment)

Main Effects; Intervention v Control, Hedges g effect sizes & 95% CIs
Please see Appendix E3 for full details on these multilevel models.

Year Group / Outcome	No of Schools	Number of Pupils	Effect size (Hedges g)	95% Confidence Intervals (CI)	Comment
Y7 (PiM12)	49	1,844	0.00	-0.11 to +0.11	Not Significant
Y8 (PiM13)	52	1,761	+0.02	-0.16 to +0.19	Not Significant
Y9 (PiM13)	45	1,648	0.00	-0.21 to +0.21	Not Significant
Y7to9 (PiM12to14)	56	5,253	-0.01	-0.10 to +0.09	Not Significant

Table 32 PiM Outcome Subscale 1 - (MR related)

All regions combined (On Treatment)
Main Effects; Intervention v Control, Hedges g effect sizes & 95% CIs

Year Group / Outcome	No of Schools	Number of Pupils	Effect size (Hedges g)	95% Confidence Intervals (CI)	Comment
Y7 (PiM12)	49	1,844	+0.02	-0.10 to +0.15	Not Significant
Y8 (PiM13)	52	1,761	+0.08	-0.12 to +0.28	Not Significant
Y9 (PiM13)	45	1,648	-0.04	-0.24 to +0.17	Not Significant

Table 33 PiM Outcome Subscale 2 - (MR related & weakly / strongly connected)

All regions combined (On Treatment)
Main Effects; Intervention v Control, Hedges g effect sizes & 95% CIs

Year Group / Outcome	No of Schools	Number of Pupils	Effect size (Hedges g)	95% Confidence Intervals (CI)	Comment
Y7 (PiM12)	49	1,844	+0.01	-0.10 to +0.12	Not Significant
Y8 (PiM13)	52	1,761	+0.09	-0.10 to +0.28	Not Significant
Y9 (PiM13)	45	1,648	0.00	-0.16 to +0.15	Not Significant

Table 34 PiM Outcome Subscale 3 - (MR related & strongly connected)

All regions combined (On treatment)
Main Effects; Intervention v Control, Hedges g effect sizes & 95% CIs

Year Group / Outcome	No of Schools	Number of Pupils	Effect size (Hedges g)	95% Confidence Intervals (CI)	Comment
Y8 (PiM13)	52	1,761	+0.12	-0.06 to +0.29	Not Significant
Y9 (PiM13)	45	1,648	-0.01	-0.14 to +0.12	Not Significant

7.7 Within Region Analyses

For the main intention to treat analyses reported in sections 7.3 and 7.4 above, the models were replicated within the three geographical regions. The effect sizes for these analyses are summarised in Table 35.

The first (all areas) row of Table 35 is a summary of Table 22 above. Rows 2 to 4 provide a similar summary but within the three geographical regions involved in the MRP trial.

In terms of negative and positive effect sizes, Table 35 displays a degree of geographical variation; ranging from +0.2 in Y8 region 1 to -0.3 in Y8 region 2. Nearly all of the effect size statistics shown in Table 38 are not statistically significant, further emphasising the 'no impact' pupil attainment conclusion.

Table 35 Summary of effect sizes across year groups and regions (Overall PiM Score)

1.1.1.1.2.2 Intention to Treat Analyses

		Year Group			
		Y7	Y8	Y9	Y7 to Y9
Geography	ALL Areas	-0.02	-0.08	-0.11	-0.07
	Region 1	+0.08	+0.20	-0.19	+0.03
	Region 2	-0.10	**- 0.34***	-0.13	**- 0.16***
	Region 3	-0.10	+0.08	+0.17	+0.01

** - g is statistically significant at the 5% level.*

Table 35 does contain one contradiction to this 'no impact' conclusion; the Y8 sample taking the PiM13 test in region 2. In this instance, a small but statistically significant negative effect size is observed (g= -0.34). This Y8 finding is also picked up for the combined Y7 toY9 analyses for region 2 (g= -0.16). Given the wealth of evidence from all other models that leads to the 'no impact' conclusion, this statistically significant finding might be discounted as a chance finding. However, before making this conclusion the Y8 sample from Region 2 was scrutinised.

Table 36 summarises effect sizes from eight multilevel models just for the Y8 subsample in region 2. The models relate to the four PiM13 outcome measures. Two models for each PiM13 outcome measures are summarised, first the 'intention to treat' analysis and second the 'on treatment' analysis.

Table 36 A focus on Y8 in Region 2

All PiM13 based outcome measures
Intention to treat analysis & on treatment analysis
Main Effects; Intervention v Control, Hedges g effect sizes & 95% CIs

	Y8 (PiM13) Outcome Measure	Effect size (Hedges g)	95% Confidence Intervals (CI)	Comment
Intention to treat	PiM13 Overall	**-0.34***	**-0.63 to -0.05**	Significant negative impact
	PiM13 (MR related)	**-0.61***	**-0.89 to -0.32**	Significant negative impact
	PiM13 (MR related & weakly / strongly connected)	**-0.56***	**-0.83 to -0.30**	Significant negative impact
	PiM13 (MR related & strongly connected)	**-0.53***	**-0.78 to -0.28**	Significant negative impact
On treatment	PiM13 Overall	**-0.50***	**-0.75 to -0.26**	Significant negative impact
	PiM13 (MR related)	**-0.57***	**-0.85 to -0.29**	Significant negative impact
	PiM13 (MR related & weakly / strongly connected)	**-0.57***	**-0.84 to -0.29**	Significant negative impact
	PiM13 (MR related & strongly connected)	**-0.52***	**-0.78 to -0.26**	Significant negative impact

In all cases, the difference is statistically significant - the on treatment analyses does not result in reducing the size of this difference.

However, given the unexpected variance structure (see Section 5.3), the statistical power of these findings needs to be looked at more closely. Table 37 provides some summary details for the region 2, Y8 subsample that will be used to assess the statistical power of the above findings.

Table 37 A focus on Y8 in Region 2

All PiM13 based outcome measures
Intention to treat analysis & on treatment analysis
Main Effects; Intervention v Control, Hedges g effect sizes & 95% CIs

	Intention to Treat Analysis	On Treatment Analysis
Number of pupils	591	478
Number of classes	33	27
Number of schools	17	16
Mean pupils per class	17.9	17.7
Mean classes per school	1.9	1.7
R^2 KS2 maths v overall PiM13	0.71	0.77
School level ICC	46%	57%
Class level ICC	32%	20%

Using the above details, to achieve a statistical power of 80% the minimum detectable effect size (MDES) would need to be 0.8 or higher for the intention to treat analyses and 0.7 or higher for the on treatment analyses. The effect sizes shown in Table 31 are all below these MDES estimates and so will all have a statistical power lower than 80%.

To look closer, the focus is just on the overall PiM13 outcome measure. This was observed to have an effect size of -0.34 for the intention to treat analysis and -0.50 for the on treatment analysis.

For the intention to treat analyses, an effect size of -0.34 is calculated as having a statistical power of just 22%. For the on treatment analyses, an effect size of -0.50 is calculated as having a statistical power of 49%.

To summarise, a statistically significant difference between the intervention and control group was observed in Y8, region 2; the intervention group attained significantly lower on the PiM outcome compared with the control group. However, once taking account of the sample size and variance structure of the data within the Y8 region 2 subsample, these statistically significant differences were found to statistically be notably underpowered. Therefore, whilst the observed differences are present within the Y8 region 2 subsample, the evidence is not strong enough to conclude that they genuinely reflect a true difference between the control and intervention group samples. Further, if the Y8 region 2 subsample findings and statistical power issue are considered alongside all of the other findings within the impact analyses, the conclusion all points towards the same thing; that no evidence was found that the MRP teacher PD intervention had any impact on pupil level attainment in the nine month period of the study.

7.8. Summary of impact analysis

In summary, across all models and all three year groups, once the data structure, KS2 maths attainment and other statistically significant pupil and school level factors are controlled for, the difference between the intervention and control group samples is not statistically significant - affirming the 'no impact' conclusion.

Table 38 provides a summary of the intention to treat and on treatment analyses using effect size statistics. Effect sizes observed within the samples are shown to range from weakly positive (g=+0.12) through zero and on to weakly negative (g=-0.15) but none were found to be statistically significant.

It should be noted that this 'no impact' conclusion relates specifically to the PiM outcome measures (and the three subscales). The MRP intervention was a teacher PD intervention and the trial was looking for impact at the pupil level within a six month time frame. So the 'no impact' conclusion should be considered alongside the trials relatively short time scale and indirect (teacher) focus.

Table 38 Summarising effect sizes and directions for main analyses

The MRP clustered RCT multilevel analyses

Comparing intention to treat and on treatment findings

		Y7	Y8	Y9	Y7 - Y9 combined
PiMs Overall (General Maths Attainment)	Intention to treat	-0.02	-0.08	-0.11	-0.07
	On treatment	0.00	+0.02	0.00	-0.01
PiMs Subscale 1 (items with some relationship with multiplicative reasoning)	Intention to treat	0.00	-0.02	-0.15	n/a
	On treatment	+0.02	+0.08	-0.04	n/a
PiMs Subscale 2 (items with some relationship with multiplicative reasoning that are weakly or strongly connected to topics in the MRP project)	Intention to treat	0.00	+0.01	-0.09	n/a
	On treatment	+0.01	+0.09	0.00	n/a
PiMs Subscale 3 (items with some relationship with multiplicative reasoning that are weakly or strongly connected to topics in the MRP project)	Intention to treat	n/a	+0.02	-0.08	n/a
	On treatment	n/a	+0.12	-0.01	n/a

7.9 Control school approach

Control schools were given no explicit steer from the research team to alter anything about their approach to PD, curriculum materials or mathematic pedagogy practice once it was confirmed they would not be intervention schools. However, in an attempt to understand if there were any particular issues that might have influenced pupil performance in control schools, a very simple online survey was sent to all the nominated control school CTs from school still remaining in the study (n=54) post the test in July 2014.

Seventeen CTs completed the survey, from 15 schools (in two schools both CTs completed the survey) this represents a response rate of only 31% in terms of overall control school CTs; however in terms of school coverage over half (56%) are represented (ie at least one CT filled in the survey). The table below outlines the three core closed questions and teacher responses to them.

Table 39 Data on control school approaches

Closed questions from the survey	Number of yes responses (%)	Number of no responses (%)
1) Are there any particular circumstances in your school or department that might have influenced test outcomes of pupils who were tested as part of the Multiplicative Reasoning Project?	3 (18%)	14 (82%)
2) Did you make multiplicative reasoning a particular focus of your KS3 teaching in 2013/14?	1 (6%)	16 (94%)
3) Has the mathematics department been engaged in any unusual or new form of professional development in 2013/14, for example Lesson Study?	7 (41%)	10 (59%)

Under a fifth of teachers felt there had been a particular circumstances at school or departmental level that might have influenced the test outcomes of pupils tested, and two of the teachers were from the same school. The open comments linked to question 1 reveal that the school with two teacher responses, there had been several changes of teacher throughout the year due to staffing issues and a trip out on the day of the test meant very low numbers of year 9 pupils undertaking the test. In the case of the other school, the respondent pointed out that their year 8 groups consisted of low ability pupils which they felt accounted for their results being much lower than the national average.

Only one teacher from a control school felt that MR had been a particular focus of their KS3 teaching during the year of the RCT, '*Each term the scheme of learning is themed to focus on one area of multiplicative reasoning*'.

However, the survey results indicate that some control school mathematics departments are quite likely to have engaged in an 'unusual or new form of PD' during the course of the RCT, with over 40% of teacher's feeling their department had. The following examples are provided by teachers; participation in lesson study, peer observations, new innovative cross curricular approaches for example '*The students…compose music in their maths lessons with a focus on keeping the same number of beats in a bar using fractions*', involvement in a separate prominent piece of national research around mathematical problem solving, Active Learning and Study skills.

Although the numbers involved are low and the survey basic, the results do offer an insight into how the majority of control schools do appear to be largely operating a business as usual approach to teaching maths, very rarely specifically focusing on MR; offering little to suggest any widespread or fundamental problems with the testing itself but evidencing as you might typically expect some new approaches to PD.

8. School narratives

In this section we draw on the 9 case studies to provide brief narratives of how the project was implemented. Further detail is given in Sections 8.2, 8.3 and 8.4 in relation to specific issues.

Case study A13

- Located in a semi-rural market town
- Larger than average middle school (Y5-Y8)
- Predominately White British pupils - low EAL and lower than average SEN and FSM.
- Attainment is above the national average with pupils making 'outstanding' progress.
- Outstanding Ofsted (Maths department selected as best practice visit)

The school originally became involved via CT1 seeing the call on the NCETM website. The MRP was seen to fit in with the school's underlying ethos and pedagogical foci centred on pupils understanding the concepts underpinning maths.

> *We trust the teachers, but we have been aware of the need to develop the type of teaching approaches that engender understanding, rather than just mechanically solving problems. Anything that could help that, and the idea that if multiplicative reasoning unlocks attainment, then we wanted to be involved (A13, CT1).*

CT2 was keen to be involved to aid own PD at KS3 as well as for upper KS2:

> *I was equally enthusiastic, as much for my own professional development at Key Stage 3, but also bringing the experience of Key Stage 2 knowledge and how we approach maths at Key Stage 2 further up. I was very enthusiastic about joining in and getting involved (S13, CT2)*

The headteacher reported greater confidence that the project was delivered through a 'respected body' such as NCETM which left him more reassured about school involvement compared to other potential PD.

The CTs had used all the materials with at least some classes and had also adapted some for use in KS2. Materials were shared with the department. The CTs involved, one of whom is the HoD, were positive about the benefits of the project and the materials and enthusiastic about incorporating them in schemes of work next year. The headteacher was more circumspect being concerned that future change should happen gradually (emphasised how MR remained one part of a very broad subject), and he believed that the presentation of some of the materials needed to be improved.

Case study A43

- Located in suburban area
- Larger than average secondary school
- The majority of pupils White British
- Pupil premium eligibility significantly under national average
- Above national average proportion of SEN pupils
- Good Ofsted

The school joined the project following a PD lead approaching one of the CTs in the project. At this school three teachers participated in the various TIME events. The original plan was for CT1 and CT2 to attend all the lessons but for personal reasons CT1 was unable to make the first time meeting. This prompted CT1 to approach CT3 as to whether they would like to be involved instead. Having participated in the first TIME team event and enjoying it, the decision was taken for CT3 to attend subsequent TIME team events instead, then feedback and work as a three.

The CTs stated that the MRP's approach broadly aligned with the school's existing approach to teaching mathematics in that they try to promote and encourage discussion and questioning. The HoD was said to have been supportive but nevertheless CT2/3 were very glad there were three staff members involved as it mitigated the risk.

> *Yes, I really do think it shared the risk. Also there are three of you to go up to the Head of Department. (A43, CT3)*

The school was also said to be supportive, although on one occasion cover was not arranged and so teachers could not attend the TIME event.

> *Just get on with it, do whatever you want to kind of thing, more than happy. (A43, CT3)*

The CTs valued the professional development experience and were positive about the curriculum materials. They saw them as being useful for a range of different pupils. The CTs were more critical about the delivery and logistics of the programme finding a lack of overview at the start problematic and wanted more discussion time with other teachers.

CTs generally reported being supported by the senior leader team (the cover incident apart) - but they did report being mindful of not being observed doing this content as they were worried that it would not necessarily showcase progress due to the predominance of verbal as opposed to written outputs.

Despite reporting that the HoD was supportive, CTs still hinted at an underlying unease that they had to a large extent left the 'normal' curriculum behind and that the accountability system of having to clearly demonstrate progress remained. This combined with the fact that MR lessons tended to focus on a lot of verbal outputs rather than written outputs, meant that CTs were mindful about being observed teaching an MR lesson for a formal lesson observation.

Also it was a slightly weird thing to be doing. You completely came out of the normal curriculum map, as we have, and you were teaching something completely different. So to do that as a standalone teacher within a school, I think you'd really have to hold your nerve that you were doing the right thing and stick with it. The fact that everybody else is doing something completely different (A43: CT3)

The CTs were positive about the overall approach of the project but quite critical about aspects of the delivery and the logistics.

Case Study A46

- Located in suburban area
- Larger than average secondary school with a sixth form
- Has Leading Edge/training school/Maths Hub status
- Number of EAL/minority ethnic background pupils is low in comparison to national averages
- Pupil premium eligibility well under national average
- Above national average proportion of SEN pupils
- Outstanding Ofsted

The school has a proactive and supportive approach to PD. The headteacher has a clear vision and rationale for PD of their staff, for example many staff were undertaking flexible Masters degrees at a nearby university. The headteacher undertook a doctorate about how children learn in science and was keenly involved in a big research project run by a local university focused on children's leaning and science. This has led to a concern about misconceptions and pupil's conceptual structures in mathematics as well. The department had been involved in the ICCAMs project, and were positive about the materials produced as part of that in relation to MR and so responded to an NCETM advert about the project. The Key stage 3 coordinator and a recently qualified teacher were selected to be involved.

The school pedagogical approach is to emphasise understanding rather than overly rapid coverage of curriculum. The school is also involved in a Y6/Y7 transition project that has university involvement as well as another project about A-level mathematics. The school intends to conduct an Y6/Y7 multiplicative reasoning project in the future.

The department scheme of work provides a structure in terms of content, but approaches to teaching it are decided by the teacher but with collaboration and sharing of practice with the department. The approach to teaching mathematics aligns well with the MRP activities.

As a department we look for conceptual understanding through group work, through student talk and through developing ideas, rather than front of classroom delivery. Although obviously that does happen because of course it will in a maths department, but we promote a lot of investigative tasks in our department meetings. So we'll explore investigative tasks to help support the department deliver them. (A46, HoD)

The HoD believed that the MRP materials could potentially support the conceptual understanding of low attainers. There are plans for materials to inform the departmental development plan next year.

In this school, school leadership was highly supportive and invested. PD is valued and thoroughly established. The MRP fitted squarely with an existing trajectory. A46 is a teaching school and was very keen to be involved, has an outstanding Ofsted judgement and is now a Maths Hub school with wider responsibilities in the area. All these factors combined to create a culture within the school that was much more conducive to risk taking and embracing new initiatives such as MRP that are considered to be potentially beneficial.

Plans were already afoot for incorporating MRP materials into the schemes of work next year, doing more Lesson Study (experienced and inexperienced pairing with similar attainment classes) and with a new hubs based initiative.

Case study B1

- Located on the outskirts of a market town in a rural county
- Small community 11-16 school (roll < 500)
- The proportion of pupils with SEN is below average
- Pupil premium eligibility is in line with the national average
- Good Ofsted
- Improving school with attainment rising over the past three years

There are three teachers and two support assistants in the mathematics department. The Head of Department was also CT1. CT2 was not a maths specialist but had been teaching maths for several years. The department (teachers and TAs) meet every 3 or 4 weeks for 'learning meetings' and have a common approach to teaching:

> *Child-centred learning, the having learning leaders within the group, children talking maths, challenging questions, putting things into context, problem-solving. (B1, CT1/HoD).*

HoD/CT1 saw the project advertised on the NCETM website and volunteered. She encouraged CT2 to become involved. The headteacher felt that the project 'couldn't have come at a better time' as the school were involved in various professional development initiatives, including action research. He was overwhelmingly positive about the department's involvement in the project, as were all the teachers in the department. The enthusiasm for the project was evident from all the interviews:

> *I've found it really, really interesting. The ideas have been great, and we've just been trying to apply it to the department really. We've all got quite excited about it – it's been really good. (B1, CT2).*

The department had plans to look at the materials and to build them in to schemes of work:

There's loads we haven't really had time to look at yet. So it will be good to sit down and at some point have a look at it and say, 'Right, let's have a look at this now. We've done that, let's try that.(B1, CT1/HoD).

Case study B2

- Located in a small town in a semi-rural setting with good transport links to major cities
- Large 11-18 academy school
- The number of students eligible for free school meals is below the national average
- Proportion of children from minority groups is low
- School judged good by Ofsted
- Proportion of students achieving 5+ A*-C GCSE was above the national average in 2013

The HoD spoke positively about mathematics in the school but felt that the involvement in the project had pushed them into action to improve teaching further, getting the department to work together to discuss student learning:

I think what this has kick-started almost is that debate on what makes a good lesson and trying to reflect on what the pupils have learnt, where – if anywhere – there's been that first misconception. After that first misconception everything is plummeting and it's impossible to get them back when you've lost them. I think the best thing about it is it's genuinely kick-started our ambition to share good practice and do it, not just talk about it. (B2, HoD).

The HoD believed that participation in the project had refocused the department's attention on students' learning. CT2 saw the advertisement for the project and 'thought that sounds really good and just applied for it.' (B2, CT2). She wasn't sure what the project was about and acknowledges that they took a risk with it. CT1 welcomed the opportunity to go on the course with her:

The Head of Department was quite keen for us both to go, you know, so that we could do the sharing ideas and all that sort of thing. I think he thought seeing as we both had positions of responsibility then it would make sense for us to go. (B2, CT1).

The HoD noted that *'anything that might improve our mathematics teaching is worthwhile doing and the head teacher gave her support'. (B2, HoD)*

The bar method and ratio tables were seen as powerful models with evidence that they were beginning to impact on student understanding. The teachers involved were very positive about the quality of professional development. The intention is to incorporate the lessons into their schemes of work. The core teachers were encouraged to lead professional development in the school, building on the model they had experienced.

Case study B36

- School situated in a rural county
- Smaller than average recently converted 11-16 academy school
- The proportion of SEN pupils and those with disabilities that need support through school action is significantly below average.
- The percentage of students eligible for the pupil premium/entitle to FSM is below the national average
- Good Ofsted judgement
- Improving results although attainment in mathematics was below average and there was said to be insufficient challenge in mathematics lessons

The HoD (in post for one year) described the department as 'a changing department' with 'see-sawing results', one that had been quite traditional:

> *Very strong in terms of KS4 exam preparation […] not so good in terms of developing understanding at KS3 (B36, HoD).*

Changes started 3-4 years previously with a focus on 'trying to teach things better the first time' and developed this year aiming for 'real understanding', moving away from textbooks and worksheets to using 'models and images' with 'a lot more group work, a lot more discussion work, a lot less didactic teaching' (B36, HoD). He reported that students were happier with the approach but that staff had taken it on to varying degrees. HoD cited a critical Ofsted inspection as giving them licence to make changes in the department; they have done this by encouraging staff to try things out….'pushing higher order thinking skills rather than more of the same' (B36, HoD).

The department was working to share ideas, use more practical activities and resources. HoD saw the biggest issue as 'trying to convince students who arrive at the school convinced they can't do it'. He felt that 'they arrive with tricks to pass exams […] and methods they learn without a core understanding' (B36, HoD). There was said to be a culture of low aspirations in the area.

The maths department set students for maths and historically there had been a view that 'we can teach them all [in a set] the same thing' (B36, HoD) without acknowledging differences within sets. The department had staffing issues, losing two members of staff (1 f/t & 1 pt/) from a team of 6 FTE. The p/t teacher was replaced (though with a non-specialist), the other loss was managed through group changes and addressing timetabling issues. The changes included collapsing groups, forming 5 mixed-ability groups in Year 9 that students moved through 5 different activities.

HoD reports that involvement was through CT1 who brought the MRP project to his attention, 'it seemed to fit with where we want to go as a school' and was something CT1 was keen to do and HoD was keen to support it. HoD selected CT2, a 'really good teacher' (B36, HoD), a non-specialist who was working in the mathematics department as he saw it as 'an opportunity for her to get to grips with some of the depth in the maths'(B36, HoD)

According to the HoD, teaching approaches promoted by the programme were similar to departmental ones. The discussions about the MRP had been 'more on a whole-class level' (B36, CT1). The HoD reported challenges of being involved in the project such as the time involved, particularly the travelling. This was more of an issue for the non-specialist teacher as she did not teach full-time in the mathematics department and her absence impacted on other teaching.

There had been limited impact on other members of department as there were limited opportunities to share, though a new member of staff (3 weeks in post) had immediately picked up some of the resources. CT1 had led two sessions with the department though HoD notes fewer training opportunities than would have been ideal due to other pressures. CT1 was leaving the school at the end of the year, and CT2 having no mathematics teaching in the next academic year, would inevitably impact upon future plans.

Despite the loss of both CTs to the mathematics department at the end of the academic year, the HoD was positive about the benefits of the programme: 'some really good stuff there that going forward will be more and more useful' (B36, HoD). He felt that 'in a sense it has been a gathering exercise for us' (B36, HoD) and was looking to 'revamp' some of the schemes of work, building in some of the MRP materials to form a 'foundation block'. HoD notes that materials have 'worked better with some of the lower ability groups' (B36, HoD) as they are more willing to try the new approaches; the more able students try to use methods they already know despite not fully understanding them. HoD plans to use materials/approaches in Key Stage 3, at the start of topics, and informally in Key Stage 4.

Case Study C8

- Located on the edge of a large city
- Large 11-18 academy convertor school that leads a small academy chain
- Below average FSM and ESL.
- Teaching school and MESH host
- Largely white British ethnicity with low SEN
- School judged 'Outstanding' by Ofsted
- Attainment is in the highest quintile but in second quintile for similar schools and third quintile for mathematics in comparison with similar schools.

The school is a teaching school and was a MESH host, so the TIME team events took place on site. The school also provided one of the TIME team PD leads. The mathematics department is involved in other initiatives, for example, one of the CTs (CT2) is involved in a University led lesson study project. The school was invited to be a MESH and provide a PD lead due to pervious involvement with the NCETM.

CT1 was the Key Stage 3 coordinator, however she was part time and only teaching one Y7 class. She was selective about which materials she used. CT2 was not initially nominated. She had recently qualified and was also a participant with a lesson study project and previously a project on problem solving. This teacher made greater use of the materials. For one of the two teachers the teaching approach promoted by the project was new, for the other the approach was more familiar. The bar model was seen as effective by both teachers and pupils.

Lesson study was not promoted beyond the first use in this TIME team and so CT1 and CT2 engaged in one lesson study. However, CT2 stated that this was more of a reciprocal observation than a lesson study. Lessons led to greater discussion between pupils than normal lessons.

The HoD said lessons fitted with the way they like to teach anyway (but note that for CT1 there was more of a sense of change in practice and innovation). Due to the change in teacher, classes taught by CT2 experienced the MRP approach but were not tested.

The school is now leading a Maths Hub, they intend to offer the MRP materials to other schools, but will adapt the project, probably using fewer materials. They may develop a similar project related to algebra.

Case study C37

- Located in an urban location
- Large 11-18 school (>1500 pupils on roll)
- The school is a sponsored academy and part of a large academy chain
- Number of pupils eligible pupil premium is average as is those eligible for FSM.
- Average proportion of students from minority ethnic backgrounds, and a similar proportion speaks EAL
- School judged 'good' by Ofsted

There are 15 teachers in the department, with several NQTs and part-time staff; the current HoD has been in post for four years, with CT1 newly promoted to second in department in 2013. The senior director of mathematics for the academy chain saw the project advertised and recommended that CT1 consider it as a professional development opportunity that would support her in gaining promotion to third in the department. CT2 was selected as she taught in KS3. CT2 had limited involvement, attending only the first two TIME events (unit 0 and 1) before being on sick leave from February. There was no other teacher able to take over the project work and this appeared to limit the extent to which the project was implemented:

> *So we haven’t seen that much of it because not much of it has gone on. (C37, HoD)*

CT1 & 2 did one lesson study together.

There is a common approach to many issues across the academy chain and within the department which has a collaborative approach to developing consistent lessons. The department does use open ended tasks but does not usually extend tasks over more than one lesson:

[…] We don't like things to take more than one lesson. We don't like tasks to be – not open-ended, because we like that – but we don't like it to be that if somebody misses the start they are then behind for the next lesson. So they'll [the MRP materials] probably need a lot of adjusting for us to include them in our scheme of work. (C37, HoD)

This common approach was apparent as CT1 discussed professional development run by the academy chain: both the HoD and CT1 reported that pedagogical approaches used in the MRP project were not that different to those currently used in the department. However it seemed that their usual approach was more structured and they found the more flexible approaches underlying the materials challenging.

From a working together point of view then it's quite similar. The students are all used to that, and they are used to having quite a visual approach. We try to appeal to all the different kinds of learning. I think from a free rein, this is your task; you're going to do it for an hour, that's something completely different. We don't normally let them go for that long on the same thing. I think to start with they did struggle with that amount of concentration, so to try and keep up the same thing for an hour. I think the teacher also struggled with that because they're used to having more input than was needed. (C37, HoD).

CT1 notes that seeing how the lessons might be used in the scheme of work was initially challenging.

I couldn't at the time, but now with the new specification coming in. I'm seeing more links between the multiplicative reasoning and the new specification that's coming in for the current Year 9s. (C37, CT1).

However, The HoD believed that the materials would be of limited use, mainly because the approach and timings didn't match the department's preferred way of working.

We will look through them, because we're currently writing a new scheme of work for the new specification, so we'll look through them and see what we can use and what we can't. It will be limited because of the time. (C37, HoD)

Case study C39

- Suburban location
- Large 11-19 school (>1000 pupils on roll)
- The proportions of students who are disabled or with SENs lower than average
- The percentage of pupils eligible for FSM below average
- The proportions of students from minority ethnic groups, and EAL are much lower than the national average
- Second quintile of national performance but below average for similar schools
- School judged 'good' by Ofsted

The initiative to participate came from CT2 a non-specialist teacher of mathematics who saw the advert on the NCETM website. Initially it was planned that the head of Key Stage 3 mathematics would participate but because the Key Stage 3 lead was planning to leave, CT 1 was substituted as an experienced mathematics specialist.

CT 2 - a non-specialist mathematics teacher who was trained as dance teacher, - had done a professional development course for non-specialist teachers at a university. The HoD supported the teachers' application but needed to convince senior leaders to agree to release teachers. The two teachers spoke at department meetings but have largely undertaken the project independently of the rest of the department.

> *I think because there are a lot of staff here who are nervous about losing control of the class I think, and letting them make the mistakes. We've been trying to encourage them, haven't we, particularly with the NQT. Trying to get them to let them make the mistakes and let them do it. I think they're getting there. I think some people just have their way of teaching and this isn't how they would normally teach a lesson and they were nervous to try it. (C39, CT1)*

The project appeared to have had significant impacts on the CTs' practices and thinking. They made extensive use of materials, all of which were used with at least some pupils. However, the two CTs were quite isolated and going against the grain. The project was accepted rather than really supported by the HoD and the school. Due to this relationship to the department it is not clear if materials will be incorporated into schemes of work next year.

9. Security of findings

9.1 Introduction

In this section we discuss the security of the findings of the trials and then go on to discuss possible explanations for the finding and the implications of this.

9.2 Dimensions of security

In discussing the security of the findings we draw on the approach developed by the EEF[35].

The EEF have adopted a 0-5 scale for classifying findings of randomised control trials. 0 represents a trial that adds little to the evidence base, 5 representing the strongest possible evidence for a single trial.

The EEF propose five aspects to consider - Design, Power, Attrition, Balance and Threats to Validity. Each of these aspects can be ranked on a scale of 0-5, so developing an overall description of the security of the trial[36].

9.2.1 Design

The RCT design met consort standards and was clear and fair. There are some known issues with the design. So for example, two of the schools in one TIME team applied to take part in the project even though the PD leads were recruited from them. They were then included in the randomisation. They could not reasonably have been allocated to a control group. Fortunately, they were randomly allocated to be intervention schools rather than controls, which still potentially impacts on the security. In addition, in one other TIME team there was a school allocated to the control group that was part of the teaching school Alliance led by the school that was hosting the TIME team and providing PD leads. We also know there was considerable drop out in terms of recruitment from initial interest. The overall recruitment to the trial resulted in a sample slightly above the national average (this will be discussed further in relation to validity). So overall we judge the design to fall short of the highest level possible - so we score it as 4.

However, it is important to note that the design of the trial was premised on a project in which the focus would be on the teaching of MR in a specific way. In the event, as is discussed in implementation/process evaluation section in the final report the project differed from this with a greater emphasis on wider professional development goals. Given the nature of the project as it evolved, if a similar project was to be repeated then an RCT that takes the teacher as unit

---

[35] EEF (2014) Classifying the security of EEF findings: Note this was developed in collaboration with Stephen Gorrard and Steve Higgins.

[36] Note that the security of trials is assessed not only by the evaluator but at least two independent peer reviewers of EEF reports. Our self-assessment should be treated with caution.

of evaluation would be appropriate, this would introduce other questions about balancing the sample at teacher level on criteria that were not included in this trial.

9.2.1 Statistical power and sensitivity

The statistical power relates to the probability of detecting a genuine effect. There are two sources of statistical 'error'; false negative (Type I)[37] and false positive (Type II)[38] errors. In this trial, the type I error were fixed to be 0.05 (5%) or lower and the type II error to be 0.20 or lower. This resulted in a trial that had a predicted statistical power of 0.8 (or 80%) which is the advised CONSORT standard. In adopting an 80% statistical power, a minimum detectable effect size (MDES) of '0.2' was estimated based on the research design and sample sizes.

In summary, the trial was designed to be able to detect a minimum detectable effect size of 0.2 or greater as being statistically significant with a statistical power of 80%.

9.2.2. Attrition

Attrition occurred at three levels: school, teacher and pupil. This is detailed elsewhere in the report.

However, for the purposes of classifying the security RCT as a headline we consider intention to treat. And so the overall attrition is a little over 20%. Note that this does not take into account two control schools who dropped out before the commencement of the trial - this issue will be considered under validity.

9.2.3 Balance

Balance refers to the balance of the sample. Note that here the issue is of balance in relation to pupils on observable characteristics. This is limited. So for example, qualities of the teachers themselves were not observed beyond the numbers or pupils taught by two different sorts of teachers - the experienced and less experienced teachers.

To recap at the point of randomisation, we stratified the sample using a relatively crude measure of the overall attainment of the school - above or below national average on 5 A*-C GCSEs. Analysis of NPD data allows a finer grained picture of balance to be obtained. This shows that the control and intervention groups were unbalanced in some important respects as discussed in Section. 7.2.2 Further, we have identified an issue with school type.

Overall, we consider the score for balance to be 3. It should be noted that one difficulty with the issue of balance is that small imbalances may be highly significant in some contexts whilst relatively large imbalances may not matter very much at all.

---

[37] This is when the null hypothesis (no statistically significant difference between the control group and intervention group) is accepted / maintained when this is not actually the case. In this design, this is fixed to have a probability of 0.05 (5%) or less.

[38] This is when the alternative hypothesis

9.2.4 Validity

It is useful to consider two aspects to validity - internal and external validity.

Internal validity - in randomised control trials, fidelity is important to internal validity. Where there is high fidelity all pupils have similar learning experiences that are facilitated by teachers who have had similar professional development experiences. A casual inference is then possible that it is the intervention that is the causal explanation. However, in the MRP fidelity was relatively low (as is discussed in the 'implementation' section of the main report). Therefore, it is difficult to know the extent to which different aspects of the programme were important. It is conceivable, for example, that if teachers had attended the events but had not been asked to use the materials as teaching materials this might have had positive impact. Or if they had used the materials without the PD, because of the variety of ways the project influenced the student experience, it is difficult to assess impact.

Also low fidelity means that it is not clear how much 'treatment' some pupils got. For example, in some cases the pupils tested were not the pupils of the teachers who attended the PD events. Additionally, teachers used the materials to varying extents. Note that from case study visits, it is apparent that the surveys of material use are a blunt measure as in some cases teachers used materials in a unit over a number of lessons, where others, teachers used items for a shorter amount of time.

For **external validity** it is important to consider the nature of the trial itself. Three different sorts of trials are possible - pilot, efficacy and effectiveness. Given the project was in development, and that the overall focus evolved, this was clearly a pilot. Given this the external validity is low both in terms of a similar project, and even more so in respect of other interventions focus on MR or on other aspects of the professional development. In addition there are other issues in the trial that are not covered explicitly in the EEF model, for example, whilst we endeavoured to ensure all testing was invigilated and conducted in a standard way, we know that this was not achieved in all cases. In addition two very high scoring control grammar schools did not have external invigilators - one because of illness and one because of choosing to make their own arrangements.

9.3 Overall assessment of security

The table below provides a summary of our assessment of security.

Table 9 Summary of security rating

Rating	1. Design	2. Power (MDES)	3. Attrition	4. Balance	5. Threats to validity
5	Fair and clear experimental design (RCT)	< 0.2	< 10%	Well-balanced on observables	No threats to validity
4	Fair and clear experimental design (RCT, RDD)	< 0.3	< 20%		
3	Well-matched comparison (quasi-experiment)	< 0.4	< 30%		
2	Matched comparison (quasi-experiment)	< 0.5	< 40%		
1	Comparison group with poor or no matching	< 0.6	< 50%		
0	No comparator	> 0.6	> 50%	Imbalanced on observables	Significant threats

Thus, we propose an overall security rating of 2. This would suggest that repeating the MRP in its current form may lead to a similar outcome in terms of impact. However, it does not necessarily mean that elements of the MRP could not have an impact or, if there was greater fidelity, that this might not lead to impact.

Annex A - NCETM Documentation

A1: Project Overview

National Centre for Excellence in the Teaching of Mathematics	KS3 Multiplicative Reasoning Project Overview

1. The primary purpose of the project is to design and test a development programme that will lead to improved outcomes for pupils in KS3 in the area of multiplicative reasoning.

2. As well as improved outcomes for pupils, the development programme will support further outcomes including:

- professional development for the teachers participating in the programme (both the core teachers and their departmental colleagues)
- leadership development for the teachers co-leading each local development team
- curriculum and resource development resulting in teaching plans and resources (three teaching units for each KS3 year)
- the development of practices and resources that support teachers in being evidence informed practitioners
- the establishment of Mathematics Education Strategic Hubs (MESH) led by teaching schools that can sustain the model of development in the future

3. There will be three *TIME (Teachers Improving Mathematics Education) teams* in different parts of the country. Each team will be made up of:

- 20 teachers from ten schools (one teacher from the school will have a management role at KS3 and the other will be an inexperienced or non-specialist teacher). Schools will be randomly selected to participate in the programme by the external evaluator from a pool of appropriate and interested schools
- two mathematics professional development leads (normally teachers from Teaching Schools or NCETM accredited PD Leads), who will lead the work of the development group
- a local mathematics education researcher and research assistant who will ensure teachers have access to relevant research and can use evidence that emerges from their own practice
- the MESH co-ordinator from the Teaching School, who will manage local communications, administration and partnership relationships

4. There will be a *national development team* (NDT) who will develop the framework for the teaching plans and the professional development programme. The team will comprise of:

- the PD Leads and mathematics education researcher from each of the TIME teams
- three curriculum and resource developers, who will support the development of teaching plans, associated resources, and PD materials
- the NCETM project lead and project manager (supported by the NCETM Secondary Director)
- a national mathematics education expert in the field of multiplicative reasoning

5. The NCETM will be supported by a *national advisory group* (NAG) consisting of mathematics educators, officials and research experts. The NCETM project lead and the NCETM Secondary Director will report to this group, which will be chaired by the NCETM Director.

6. Alongside the internal evaluation carried out by NCETM with the NDT and TIME teams, there will be an external evaluation of the development programme. This will use randomised control trials in assessing impact on pupils (for each TIME team, the 10 schools participating in the programme will be randomly selected from a wider group of 20 eligible schools).

7. The potential reach of the three TIME teams is as follows (assuming a model of the participating schools being on average six form entry and maths departments being six teachers):

	Number in project	KS3 pupils/teacher	Total number of pupils
Core Teachers	60	90	5400
Department Colleagues	120	90	10800

This suggests a cost of the development programme of just under £20/pupil.

8. For each school involved in the project, there will be the following outputs:

- Two core teachers will be part of a TIME team and participate in the six days of workshops during the year.
- The workshops will combine CPD activity with curriculum development. In particular, they will include work on subject knowledge for teaching; using evidence about effective teaching of multiplicative reasoning (both from research and the teachers' own classroom); and collaborative planning of the nine teaching units (three each in Y7/Y8/Y9).
- The two core teachers will then work with their department to share the learning from the workshops and to prepare for the teaching of the units.
- The two core teachers will keep a professional learning journal throughout the year. With the support of research assistants and colleagues, they will systematically collect evidence about how well the planned lessons worked and their impact on learning.

9. A possible schedule of work is shown in the table below:

May 2013	Recruitment of MESH schools, PD Leads, university experts and curriculum developers for the TIME teams Recruitment of external evaluators and agreement of NAG membership
June 2013	Recruitment of MESH schools, PD Leads, university experts and curriculum developers for the TIME teams completed Recruitment of project schools Recruitment of external evaluators and agreement of NAG membership Core team develops programme for the project schools
July 2013	School application process completed NAG (1/2 day meeting)

	Core team develops programme for the PD leads Core team develops resources for formative assessment and first taught units.
September 2013	NDT (2 day planning and development workshop: overview, formative assessment units, on-going evidence collection) Allocation of schools to control and intervention groups by RCT evaluators Pre-assessment by RCT evaluators
October 2013	TIME teams (1 day workshop in the 3 local venues: Understanding the issues, formative assessment units and collecting evidence) Departmental workshops
November 2013	Formative assessment units used with Y7/8/9 Classroom observation and evidence collection NDT (1 day planning and development workshop: Units 1 and 2)
December 2013	TIME teams (2 day workshops in 3 local venues: Unit 1) Departmental workshops
January 2014	First intervention units taught to Y7/Y8/Y9 Classroom observation and evidence collection NAG (1/2 day meeting)
February 2014	TIME teams (1 day workshop in 3 local venues: Unit 2) Departmental workshops Second units taught to Y7/Y8/Y9 Classroom observation and evidence collection
March 2014	NDT (1 day planning and development workshop: Unit 3, final evaluation, taking the project forward) TIME teams (1 day workshop in 3 local venues: Unit 3) Interim assessment by RCT evaluators
April/May 2014	Departmental workshops Third units taught to Y7/Y8/Y9 Classroom observation and evidence collection
June 2014	TIME teams (1 day workshop in the 3 local venues: evaluation & embedding) Final assessment by RCT evaluators
July 2014	NDT (1 day workshop) NAG (1/2 day meeting)

KS3 multiplicative reasoning project

NCETM pilot funded by the DFE

Subject leader guidance

project schools

Spring 2013

This document includes information on

The structure of the project

The overview of the project and its mathematical focus

Expectation for participating schools and core teachers

The subject leader role

Sharing and communication

The units

The lesson/professional development documents

Professional development

Appendices:

A timeline for the project

Overview of project

Overview of professional development outcomes

How this document can be used:

Please read though the document and discuss it with your project teachers and senior managers.

Any further enquires can be referred to Natasha Chippendale: *natasha.chippendale@tribalgroup.com*

Introduction

Thank you for agreeing to be part of this important DFE funded pilot delivered through the NCETM.

This is a professional development project focussed on pupils learning of multiplicative reasoning.

The project is a pilot and aims to see what can be learnt from bringing together groups of teachers and local university expertise in partnership to focus on improving learning in this area of mathematics in schools.

Schools involvement with the project will run until July 2014. As a pilot the support and feedback from teachers and departments will help shape the project and have a strong influence on future developments. Support and feedback from Subject leaders will also be crucial and very much valued.

This guidance is therefore aimed at giving further information on the detail of the project to subject leaders and senior managers in the intervention schools in order to best support the identified core teachers and the department to engage with the project.

Multiplicative reasoning

Many situations encountered in life, work and education involve contexts and questions which are connected by the idea of proportionality. Many seemingly different strands of the maths curriculum are connected in this way as are many questions in a typical GCSE exam paper. Recognising how different questions may be connected and involve the same maths might have considerable benefit to the teaching and learning of mathematics in this area.

Therefore the mathematical focus for the project is to look at the effect on teachers and pupils of:

- **Making connections in mathematics** where the underlining structure is multiplicative.
- **Deepening the understanding of the mathematics related to solving problems** where the underlining structure is multiplicative.

The multiplicative nature of the reasoning used to solve problems arising from proportional situations often involves an understanding and application in different ways of fractions as well as decimals, percentages, ratio and proportion. Proportionality underpins many areas across the curriculum and into KS4 with applications across other subjects.

Such a range cannot be addressed in the scale of this KS3 project. The project hopes to see what can be achieved as a starting point in allowing teachers to engage with some of the issues and teaching approaches that have proven effective in practice and backed by research. As such, the project will focus strongly on teachers' professional development in terms of the subject knowledge and pedagogy related to teaching these aspects of the curriculum. Importantly the project attempts to provide a strong foundation from which further work nationally and within departments can easily be engaged with by teachers.

The structure of delivery

- The project is arranged around 3 pathfinder hubs (MESH – mathematics education strategic hub) – Nottingham, Manchester and Portsmouth.
- Each of the project TIME (Teachers Improving Mathematics Education) teams is hosted by the MESH. As well as the 10 project/intervention schools there are 10 control schools
- The project is delivered through a series of TIME team workshops in each of the above MESH areas with two *core* teachers attending from each of the ten intervention schools
- Each TIME team workshop is led by two Professional Development Leads and one local university researcher.

Evaluation

- The project is subject to an **external evaluation** carried out by Sheffield Hallam University and this includes a test in June. The details of which is arranged between the University and the school.

- There is also an **internal evaluation** from which we will attempt to learn as much as possible from the project to shape future developments and guidance. All the participants will have an important role in contributing to this and we will hope to draw on their expertise and experiences throughout the project.

Project rationale

Conceptual understanding

- Much of this area of mathematics is recognised as difficult to learn and teach. Pupils arrive from KS2 with a wide variety of understanding and knowledge including procedural application in this area.
- The approach taken therefore will be to focus less on the direct application of procedures, but more on revealing and building on children's **underlying conceptual understanding**. Exposing misconceptions that might underpin superficially understood procedures and providing a framework for developing understanding and application of the mathematics to the solution of a wide range of problems.
- This is the backdrop to supporting pupils to secure and develop their learning in this area of mathematics as they arrive from KS2 and move through KS3 to the challenges of KS4.
- Conceptual understanding is a highlighted focus in the Ofsted framework and part of the criteria defining judgement grades in the mathematics subject specific guidance published by Ofsted.
- The new curriculum highlights the development of conceptual understanding in parallel with procedural fluency.

How will we aim to do it?

Pupils making sense of problems (through the use of visual images and models)

- The starting point is to support pupils to '**make sense'** of the problems they are tackling *(in this case proportional problems).*
- A key approach will be to support and develop pupils ability to represent such problems *(often starting from their own representations- in particular a diagram)* which are then refined to more mathematical representations. This might typically involve the use of a **bar, double number line** or **ratio table,** these having evidence in research and practice in different countries suggesting their effectiveness in supporting learning in this area.
- The use of such images to model different types of proportional problems will provide the setting from which pupils can deepen their understanding of the mathematics involved and be better able to appreciate the common structure of such problems through the models used.
- A number of particular approaches reflecting slightly different research backgrounds are developed through the teaching units and these are made explicit in the lesson commentary by the developers.
- Pupils will have opportunities to explain and justify their solutions. They will work together on refining and developing their understanding from a given starting point, which will often be a realistic scenario. There will also be collaborative activities designed expose misconceptions and challenge and refine pupils current understanding.
- All of these approaches are reflected in the project teaching units and aim to provide teachers with the subject knowledge necessary to implement a more effective framework for developing learning in this area of mathematics.

Professional development

- The project is primarily a professional development project in which paying careful attention to the impact of the teaching approaches on pupils learning is a key focus.
- The professional development will relate to subject knowledge and subject pedagogy in this area including teaching and professional practice. An overview of the professional development outcomes is included in an appendix to this document.
- The pilot also allows us to learn about the effectiveness of the professional development model being used.
- Professional development will be supported by:
  - TIME team workshops
  - Through engaging with and discussing key problems throughout the project year
  - Through engaging with the lessons *(in particular the lesson commentary)* which are set up to highlight professional development features
  - Reflecting on pupil outcomes, in particular through **assessment interviews** and **lesson study** activities.
  - Through collaborating with core teacher colleague and other colleagues in the department including the subject leader
  - Through leading or supporting department professional development sessions

The core teachers will be involved in:

- **Participating in the TIME team workshops**
- **Delivering project lessons**
  - Collaborating together to plan and teach the unit lessons to their KS3 classes
- **Further professional development activities**:
  - Gathering and **analysing evidence** of impact
  - Carrying out **assessment interviews** *(mainly in unit 0 but also later in project)*
  - Carrying out one joint **lesson study** per unit and present feedback at next TIME team workshop
- **Sharing with the department**
  - Engaging and sharing developments with the rest of the department

Subject Leader support

How you can help:

- Support the core teachers plan and teach the lessons
- Arrange time to carry out the lessons study (this may include more than the two teachers by forming a lesson study group within the department)
- Discuss the work with the core teachers and consider how developments can be shared with the department. This might include supporting the teachers to lead a session

The project timeline

- A timeline for the key project workshops, events and teaching windows for the units is given as an **appendix** to this document.
- Included in this document are suggestions as to how you may be able to support the project teachers, involve others as appropriate in the department and share key developments at department meetings

Recording and communicating

We hope to be able to gather much evidence on pupils and teachers which will be useful for the project

- Pro-formas
- Portfolios of work
- Workshops
- **TIME team community forum**

Helpful resources:

Video camera *–assessment interviews,* **digital camera** *–storing snapshots of pupils work*

Visualiser - *this is a mounted webcam that enables pupils work to be projected up 'live' and discussed with the class* ***(suggestions for cheap and very effective models are available on request)***

The units:

- There are 3 teaching units and one formative assessment unit
- Each teaching unit will have a number of lessons (up to 6) requiring around two weeks of KS3 curriculum time.
- The teaching of each unit follows a TIME team workshop. These are spread over the three terms in 2013-2014.
- The dates for the TIME team workshops and teaching of units are given in the project event timeline as an appendix to this document.

Given the rationale and approach of the project is on pupils *making sense* and solving problems *(mainly through representing problems with a set of images and models)*, the focus of the units is on developing this skill and insight rather than coverage of specific areas of the curriculum.

The units therefore are under broad headings. Of course fractions, division, multiplication, ratio, proportion and the idea of equivalences will be feature strongly with the focus on understanding their properties and how to use them in solving problems. However the aim is not to produce definitive teaching units on specific areas of the curriculum.

Broad focus of units

Unit 2

Understanding and identifying proportional contexts

Unit 3

Application to a range of proportional problems.

Unit 0 formative assessment task

Looking at our pupils understanding and reasoning

This is a very important task which sets up the project focus on pupil understanding. Many pupils carry misconceptions connected with problems related to multiplicative reasoning even when they appear to be applying a procedure correctly. Although many misconceptions may be anticipated, the way pupils think about them may be very different to any assumptions we might make about their understanding.

- The task is designed around all pupils attempting 4 core questions. Here the correct answer is not the focus, but how pupils justify why they think their approach works. It provides an opportunity to gather evidence on pupil understanding which sets up the introduction of the assessment interview and later progress study.
- This is not only useful from a diagnostic point of view to inform planning but also provides PD on questioning to probe understanding and how conceptual understanding and progress might appear and be assessed when demonstrated by pupils.
- Having analysed some of the evidence the core teachers are asked to select a small number of students to probe deeper their understanding through an interview which may be videoed (**assessment interview**).
- The task is described in detail in an annex to this document

<table>
<tr><th colspan="2">Supporting Unit 0</th></tr>
<tr><td>Supporting core teachers:
➢ Finding opportunities for selected pupils (1 or 2) to be interviewed
➢ Access to video resources and advice on use
➢ Storing of outcomes
➢ Discussion and analyse of outcomes</td><td>Wider use to support department improvement?
➢ Sharing & discussion of unit 0 outcomes with dept.
➢ Role of such formative assessment activities in sow and development of further such questions.
➢ Particular role of assessment interviews and later re-interview as evidence of progress in understanding and what this might look like.</td></tr>
</table>

The lessons:

The lessons have been written in sets of 3 (mainly two lessons in each set) with each set reflecting a particular approach or background (the Dutch *'Realistic maths education'* model is an example). The approach used in each of the 3 sets is continued in all of the units. Details of these research backgrounds informing the design of the lesson are given in the lesson booklets as part of the support for professional development.

Unit 1 - Reasoning and making sense of fractions

- Students try to make sense of fractions by using them in meaningful contexts and by making use of representations and models.
- A particular focus is the *part-whole* interpretation of fractions, though students also experience fractions in terms of sharing and measurement.
- Students compare fractions, which can lead to notions about equivalence and common denominator.

Unit 2 - Understanding and identifying proportional contexts

- Students have difficulty discriminating proportional from non-proportional situations and often apply additive strategies rather than multiplicative to proportional problems and vice versa.
- Exposing students to both types of problems and allowing them to discuss the differences will contribute to their understanding.
- There will be opportunities to further develop the use of images and models to represent a number of situations.

Unit 3 - Deepening problem solving skills through application to a range of problems

- Opportunities to further develop the use of models including the ratio table to a wider range of problems across the maths curriculum.
- These might include examples from strands other than number such as algebra and geometry.

A guiding principle:

Often in the project pupils will be asked to answer a question they may already be able to get right. The objective here however is not the answer, but the ability to reflect on and justify why an approach works. The aim is to help pupils to make sense of what they know, deepening their understanding of the maths involved and how it connects with other areas so they are able to secure more fluent and accurate application of procedures to problem solving.

The use of visual images or models can present a number of advantages to support understanding and problem solving approaches. However these need to be carefully developed with pupils so that they make sense of them and how they relate to problems rather than forced as a solution to problems **hence becoming a procedure in itself**.

Supporting Units 1,2 & 3	
Supporting core teachers: ➢ Support and discuss core teachers planning for the lessons ➢ Form lesson study group *(which can involve teachers in addition to core teachers)* to Carry out a full lesson study cycle requiring planning time and lesson observation (cover may be required in some cases) ➢ Storing of outcomes ➢ Discussion and analysis of outcomes	**Wider use to support department improvement** ➢ Share unit lessons with department for other teachers to try out with their classes ➢ Discuss outcomes at department meetings ➢ Share and discuss outcomes of lesson study

The lessons as professional development documents

Lesson presentation

The lessons are designed in a format with a specific set of headings to enable teachers to plan and teach the lesson and but also to support their professional development. In fact the booklets are referred to as professional development documents' rather than just lessons, reflecting their role in the project.

The headings allow teachers to see how the key themes of the project are promoted in the lesson, the research background and areas teachers may wish to pay careful attention to when listening to pupil responses.

The lessons are designed to have enough depth to cover a range of pupils from Y7 to Y9 with advice given on adapting the lesson accordingly. This is done bearing in mind it is the justifying and reasoning that is the objective even where procedures may have been correctly applied.

Lesson headings:

- ❑ **Lesson summary**
  The aim/purpose of the lesson, type of activities the students will be involved in, and how it might link to other lessons/pre-tasks.

- ❑ **Focus of students learning**
  This is to help teachers think more explicitly about what the key learning is. A lot of this might refer to conceptual understanding rather than procedure.

- ❑ **Lesson preparation**
  Resources and materials needed for the lesson

- ❑ **The lesson**
  Any pre tasks and commentary

  The lesson is usually set out on a single page showing the key stages of the lesson and allowing teachers to see the overall shape of the lesson. Thumbnail pictures of supporting hand-outs or resources might be included to assist in this.

- ❑ **Lesson commentary**
  On a separate page a supporting commentary is given linked to each identified stage of the lesson. The commentary might refer to the significance of activities and possible pupil and teacher responses. How progress in understanding might be recognised and encouraged. How aspects of the lesson could be adapted to challenge different groups. Any professional development points including opportunities for evidence against the key PD themes and in particular subject knowledge.

- ❑ **Adapting the lesson**
  General notes on how the work might be pitched to challenge and meet the needs of different groups of pupils across KS3. More specific advice might be given in the lesson commentary.

- ❑ **Suggestions for lesson study focus**
  These are possible 'research questions' suggested by the developer that could be used as a focus in any 'lesson study', given the opportunities provided in the lesson for gathering evidence in that area.

- ❑ **Research background to the lesson**
  Notes to support the background that informed the design of the lesson. This might refer to research etc. It might also include references to further reading.

- ❑ **Resources and hand-outs**

- Attached at the back of the lesson/professional development booklet are copies of any worksheets or hand-outs that might be used in the lesson.
-

Supporting core teachers:	Wider use to support department improvement?
➢ Discussing the significance and interpretation of the sections in order to support core teacher subject knowledge and pedagogy.	➢ Discussing the headings and their significance with the department considering if the format has a wider PD use. ➢ How this could contribute to developing key aspects of the SoW and supporting non-specialist and less experienced teachers

Some key professional development activities:

Assessment interviews

As has been stated a key focus of the project is to develop pupils conceptual understanding. So some important related questions might be: -what does this look like, how can we assess it, what might progress in conceptual understanding look like?

Much of this might be apparent as part of the ongoing assessment in the classroom but often the significance of how pupils are thinking and the nature of their understanding in a particular area or question may not be apparent and we may be making certain assumptions on this. The assessment interview may follow on from a formative assessment activity and can help to provide significant evidence to assess pupil understanding and teacher professional development.

The role of such interviews could include:

- Diagnosing particular pupils misconceptions and assessment of their understanding
- Developing the skill of the teacher in using probing questions to assess understanding and expose misconceptions.
- Help to inform subsequent teaching for those pupils
- Inform the teaching for the whole class *(where the pupils have been carefully chosen)*
- Developing insight into the key learning issues in this area
- Deepen pupils understanding through the questioning though the aim is not primarily to teach pupils in the initial part of the interview

Videoing or recording of assessment interviews (*This need not focus on faces and needs to be in line with the school policy on recording)* **where there is an opportunity brings several further advantages:**

- Often important points can be picked up when reviewing the video that were missed in the original interview
- It provides a professional development resource for the department
- It can build a portfolio of evidence

Progress studies

Where the same pupils are re-interviewed at a later stage with similar or even the same question then this may reveal progress in a pupil's conceptual understanding. This might be apparent from the nature of the reasoning and language used. Where pupils have been carefully selected this may provide further evidence of progress of the class. Again the focus is on the justifications to why answers are correct rather than the correctness of the answer alone. Insecure application of a procedure hides much valuable diagnostic information.

Comparison of this interview with the previous interview of the same pupils can also be a valuable professional development tool for discussion in department meetings in particular with regard to establishing the nature of progress in conceptual understanding.

Lessons study

A key part of the project will be what we can gain by paying careful attention to the effect on learning of carefully chosen aspects of a lesson. This is a significant part of the professional development and the expectation is that core teachers will collaborate and carry out a single lessons study per unit and share the outcomes at the next TIME workshop. Guidance to support teachers in carrying out a lesson study has been provided to them.

Assessment interviews, progress studies and lesson study cycles	
Supporting core teachers: ➢ As indicated previously	**Wider use to support department improvement?** ➢ Share and discuss assessment interviews, progress studies and lesson study outcomes at department meetings ➢ Could any of these processes contribute to the department improvement plan, could they have a roll in other areas of the curriculum? ➢ What are the implications for tracking progress, dept PD

Appendix 1KS3 Multiplicative reasoning project timeline 2013 / 2014

Date: 2013/14	Activity	Purpose	Supporting core teacher activities	Suggestions for sharing and engaging department
Nottingham: **15/16 Oct** Portsmouth**: 24/25 Oct** Manchester**: 6/8 Nov**	**TIME team workshop 1** (2 days)	• Introduction to project and its key elements • Meet TIME team professional development leads and researchers • Introduce Unit 0	• **Release both core teachers for workshop** • Pupils do core questions from the Unit 0 task explaining and justifying their approaches • Evidence of outcomes selected by the teachers - *scripts, photos, audio recording etc.* • Analysis and discussion of evidence by core teachers in relation to nature of understanding and any misconceptions • Selection of suitable students for assessment interview • Carry out video recording of selected pupils • Analyse and discuss interview • Store evidence as part of project portfolio	• Share Unit 0 resources with department for other teachers to try out with their classes • Core teachers and others share selected pupil outcomes from Unit 0 including any assessment interviews, so department can analyse and discuss • Consider role of these activities in SoW for assessing conceptual understanding • Consider role of assessment interviews as part of department assessment of conceptual understanding and to support wider teacher PD • Build department portfolio of assessment interviews
20th 0ct 2013 **to** **7th Dec 2013**	**In school:** • Deliver Unit 0 Formative assessment tasks • Carry out assessment interviews	• Unit 0: Assess pupils conceptual understanding • Reveal misconceptions • Teachers use of probing questions		
Nottingham: **6 Dec** Portsmouth: **9 Dec** Manchester: **10 Dec**	**TIME team workshop 2**	• Feedback on Unit 0 • Introduce Unit 1 lessons • Introduce lessons study activity	• Release both core teachers for workshop • Core teachers teach the Unit 1 lessons to their KS3 classes. Comments added to lesson thread in TIME team community forum. • Form lesson study group (which can involve teachers in addition to core teachers) to carry out a full lesson study cycle requiring planning time and lesson observation (cover may be required in some cases) • Upload completed lesson study pro-forma to online TIME team community forum	• Share Unit 1 lessons with department for other teachers to try out with their classes • Discuss outcomes at department meeting • Share and discuss outcomes of lesson study • Consider role of lesson study as regular element of department professional development
6th December 2013 **to 7th February 2014**	**In school:** • Teach Unit 1 lessons • Carry out lesson study cycle on identified unit 1 lesson	• Unit 1: Reasoning and making sense of fractions, in particular part/whole interpretations • PD related to lesson study focus		

Date: 2013/14	Activity	Purpose	Supporting core teacher activities	Suggestions for sharing and engaging department
Nottingham: **10 Feb** Portsmouth: **13 Feb** Manchester: **14 Feb**	**TIME team workshop 3**	• Feedback on Unit 1 lessons • Feedback on lesson study • Introduce Unit 2 lessons •	• Release both core teachers for workshop • Core teachers teach the Unit 2 lessons to their KS3 classes. Comments added to lesson thread in TIME team community forum • Form lesson study group (which can involve teachers in addition to core teachers) to Carry out a full lesson study cycle requiring planning time and lesson observation (cover may be required in some cases) • Upload completed lesson study pro-forma to online TIME team community forum	• Share Unit 2 lessons with department for other teachers to try out with their classes • Discuss outcomes at dept. meeting • Share and discuss outcomes of lesson study • Consider implications for teaching in department
10th February 2014 to 24th March 2014	**In school:** • Teach Unit 2 lessons • Carry out lesson study cycle on a Unit 2 lesson	• Unit 2: Understanding and identifying proportional contexts • PD related to lesson study focus		
Nottingham: **25 March** Portsmouth**: 2 April** Manchester: **31 March**	**TIME team workshop 4**	• Feedback on Unit 2 lessons • Feedback on lesson study • Introduce Unit 3 lessons • Introduce school project evaluation activities	• Release both core teachers for workshop • Core teachers teach the Unit 3 lessons to their KS3 classes. Comments added to lesson thread in TIME team community forum. • Form lesson study group (which can involve teachers in addition to core teachers) to Carry out a full lesson study cycle requiring planning time and lesson observation (cover may be required in some cases) • Upload completed lesson study pro-forma to online TIME team community forum. • Gather information to complete overall evaluation activities with participation and contribution from subject leader	• Share Unit 3 lessons with department for other teachers to try out with their classes • Discuss outcomes at department meeting • Share and discuss outcomes of lesson study • Consider implications for teaching in department • Discuss and complete school evaluation activities for the project
25th March 2014 to 30th May 2014	**In school:** • Teach Unit 3 lessons • Carry out lesson study cycle on a unit 3 lesson • Evaluation	• Unit 3: Application to a range of proportional problems • PD related to lesson study focus • Consider evaluation activities		
Nottingham: **26 June** Manchester: **? July**Portsmouth**: ? July**	**TIME team workshop 5**	• Feedback on Unit 3 lessons • Evaluating the lesson study process • Developing MR activities • Embedding MR in the curriculum • General evaluation	• **Release both core teachers for workshop** • Core teachers work with subject leader and the department to consider how best to use the experience and materials of the project to develop a plan to embed multiplicative reasoning (including assessment, ongoing teaching, PD and monitoring of impact) into the scheme of work. • Consider next steps for the future development work	

Overview: KS3 Multiplicative Reasoning Project

Mathematical focus:

The mathematical focus for *teachers and pupils will be:*

- **Making connections in mathematics** where the underlining structure is multiplicative.
- **Deepening the understanding of the mathematics related to solving problems** where the underlining structure is multiplicative.
- *(There will be a greater focus on conceptual understanding rather than procedural approaches)*

Project units:

There will be three teaching units and a formative assessment unit, developed to support the project. Each unit will be taught following a TIME team workshop.

Their purpose is to support teachers in applying their learning within classroom practice. Also, by using common materials, TIME teams will be more able to engage in a collective enquiry.

- **Unit 0: Autumn term**
  - *Formative assessment unit*
- **Unit 1: Autumn term**
  - Strengthening understanding of fractions
- **Unit 2: Spring term**
  - Understanding and identifying proportional contexts
- **Unit 3: Summer term**
  - Deepening problem solving skills through application to a range of problems.

Key experiences reflected in the teaching units:

- Development of **conceptual understanding**
- Use of **contexts, visual images** and **models** to scaffold understanding including the **number bar, double number line** and **ratio table**
- Different **research approaches**
- **Professional development** opportunities

Professional Development

Core Teacher activities to support professional development

• Subject knowledge • Subject pedagogy knowledge • Teaching practice • Professional practice	• Participate in TIME team workshops • Plan and teach lessons from the project units to meet the needs of the pupils • Carry out a collaborative lesson study cycle for each unit • Carry out formative assessment tasks including ‘pupil assessment interviews’ for each unit • Lead departmental sessions with colleagues relating to the project work supported by the subject leader

KS3 Multiplicative Reasoning project
Overview of teachers professional development

Outcomes : *With particular reference to multiplicative reasoning, teachers gain:*

Subject knowledge:

Teachers will have a more explicit knowledge and understanding of:

- how multiplicative structures underpin and connect areas of the curriculum
- the meanings of FDPRP and the role each plays in solving problems
- the progression in Multiplicative Reasoning in the curriculum

Subject pedagogy knowledge:

Teachers will know and understand a variety of approaches and strategies based on research and pedagogical knowledge to develop pupil conceptual understanding. Including:

- recognising the research background to some of the approaches
- how to use visual images, contexts and rich tasks to scaffold and deepen understanding
- the importance of language and how to structure activities allowing pupils to explain and talk about multiplicative reasoning.
- the use of strategies including probing questions to better assess pupils learning
- a recognition of what good progress in understanding might look like in pupil learning

Teaching Practice

Teachers will as part of their on-going practice:

- make effective use of their subject knowledge and pedagogical knowledge related to multiplicative reasoning to improve learning in the classroom.
- take account of pupil prior knowledge and the progression in multiplicative skills and knowledge when planning effective lessons
- demonstrate to pupils how MR links different areas of the maths and other subject curriculums

Professional practice:

Teachers will as part of their on-going practice:

- plan and review lessons collaboratively
- use “lesson study” and “pupil longitudinal study” approaches to deepen understanding and evaluate teaching approaches
- collaboratively discuss and work on mathematical understanding with colleagues

A3: The Multiplicative Reasoning Project

Rationale

The failure of pupils to properly grasp fundamental concepts of multiplication and division, ratio, proportion and rates of change has been documented for at least the last 30 years[39]. Pupils may learn methods and techniques but frequently misapply these to situations[40]. Whilst attempts have been made to address these issues, through National Strategies and other initiatives, there has likely been little genuine progress.[41]

Watson et al (2013) outline the existing research and recommend that "a whole-school approach is necessary to ensure that the complex ideas involved in ratio and proportional reasoning are developed coherently over time, following guidance arising from the research". This concurs with Hodgen et al (2012) findings from the Increasing Student Competence and Confidence In Algebra and Multiplicative Structures (ICCAMS) research and intervention project that developing a coherent approach and using relevant models, whilst adhering to certain pedagogic principles, resulted in a significant increase in pupils' understanding of Multiplicative Structures.

This project seeks to draw together the findings from the available research to inform and develop an effective Professional Development programme for teachers to improve pupils' understanding of multiplicative structures and to enable pupils to apply models and techniques appropriately. The classroom developments are intended to be department-wide, thus ensuring that all Key Stage 3 pupils within the school have a coherent educational experience. Structures for continuing professional development of this nature will also begin to become embedded through the establishment of professional learning networks.

Defining Multiplicative Reasoning

A variety of terms have been utilised in recent years to refer to mathematical content including: multiplication by integers, decimals and fractions, percentages, ratio, proportion, enlargement, scaling and rates of change.

The term 'Multiplicative Reasoning' is being used here to encompass the understanding of the structures within these content areas and the appropriate use of models, algorithms and techniques to solve problems related to these content areas.

---

[39] Hart, K.M. (Ed) (1981), *Children's Understanding of Mathematics*: 11-16, John Murray

[40] Watson, A., Jones, K. and Pratt, D. (2013), *Key Ideas in Teaching Mathematics: Research Based Guidance for ages 9-19*, OUP

[41] Hodgen, J., Küchemann, D., Brown, M. & Coe, R. (2009), *Secondary students' understanding of mathematics 30 years on,* BERA

Objectives

The specific objectives of this project are related to the development of pupil understanding of the nature of multiplicative relationships and the appropriate use of models and algorithms to solve related problems. However, it is also intended that the structure and implementation of the Professional Development programme should provide a working model for future developmental programmes related to other mathematical content. Additionally, it is expected that pupils who develop a better understanding of these key concepts will not only enjoy their mathematics lessons more, but will develop a more positive opinion of mathematics and have an increased likelihood to continue their mathematical studies into the post-compulsory sector.

Specific outcomes for teachers

- To develop an understanding of the general nature of the difficulties experienced by pupils when learning about multiplicative structures.
- To identify and understand the specific problems experienced by the pupils in the teachers' own schools when learning about multiplicative structures.
- To develop their own subject knowledge regarding ways in which pupil understanding can be enhanced through the use of specific models and algorithms.
- To have trialled and evaluated both generic and specific resources which can be used longer term and incorporated into schemes and plans.
- To develop department-wide policies and strategies regarding effective pedagogies to employ when creating and teaching lessons and activities related to multiplicative structures.

Specific outcomes for pupils

- To develop a deeper understanding of the nature of multiplicative structures.
- To learn to use models and algorithms appropriately to solve problems involving multiplicative structures.

Project overview

The project has several strands to it and a range of people who will be involved at each level. At the heart of the project are the TIME teams (Teachers Improving Mathematics Education). Each TIME team consists of a pair of teachers from each of 10 schools. Ideally the pairs will consist of an experienced teacher with some departmental responsibility and a less experienced or non-specialist teacher of mathematics. The pairs are responsible for leading developments in their own schools, supported by the expertise of the group leaders.

The group leading each TIME team consists of two PD leads, who might be mathematics ASTs, mathematics SLEs or mathematics consultants bringing expertise in classroom matters as well as in leading professional development sessions, plus a University researcher who will bring expertise in translating the outcomes of research to classroom practice and in supporting teachers in developing skills in collecting and evaluating classroom evidence.

The work of each TIME team is as follows:

- Phase 1: Understanding the issues
    - Introduction to the issues surrounding understanding multiplicative structures
    - Introduction to some formative assessment materials and activities
    - Introduction to ways in which to gather classroom evidence objectively
    - Ways of working with and involving others to ensure that materials are used
    - Dissemination of materials to departments
    - Use of the assessment materials in school to pinpoint the issues to be addressed
    -
- Phase 2: Intervention
    - Evaluating the classroom evidence and planning for intervention work
    - Introduction to generic teaching materials
    - Planning for use of materials and for evaluation of impact
    - Dissemination of materials to departments
    - Working with department teachers to develop effective use of materials
    - Collection of evidence of impact
-
- Phase 3: Evaluation and taking the work forward
    - Assessing the progress made
    - Developing further bespoke materials to address identified needs
    - Ensuring that effective pedagogies are embedded and continued
    - Incorporating materials and activities into work schemes and plans
    - Continuing development and work beyond the time scale of the project

The TIME team leaders are supported by the Core Development Team (CDT) which consists of the project lead plus 3 developers, chosen for their combined national expertise in understanding and researching the issues involved in Multiplicative Reasoning, developing resources and devising and leading professional development activities with both teachers and PD leaders.

The CDT, guided by a National Advisory Group, will decide upon a more precise locus and remit for the project in the early stages and will be responsible for creating resources and devising an outline plan for the TIME team professional development activities. These PD activities will be finalised by the National Development Team (NDT) which will consist of the CDT plus the TIME team leaders. This group will meet to plan the implementation of the project with the TIME teams.

Randomised Controlled

The DfE will appoint an external body to carry out an RCT. This will involve initial and final assessments plus allocation of schools to either control or intervention groups.

Key Activities Timeline (approximate)

June: Recruitment of Core and National development teams

July: School recruitment; shaping of the project locus and remit; development of formative assessment resources

August: Development of PD sessions

September: RCT allocation to groups and initial assessments.

Development of intervention resources and activities.

October: TIME teams phase 1 activities

November: Development of intervention resources and activities.

December: TIME teams phase 2 activities

March: RCT interim assessment

June: TIME teams phase 3 activities

July: RCT final assessments

Annex B - MRP materials

TIME 1 - October - November 2013

(1) Handouts

- s1 GCSE Q Booklet
- s1 Handout 1.1
- S1 Handout 1.2 Flour Scenario
- S2 KS3 Maths MR Project Overview
- S2 Overview Teacher PD
- S2 Secondary National Curriculum Mathematics
- S4 CSMS & TIMSS questions
- S4 Mistakes and Misconceptions
- S5 Q Selection Unit 0 Core Q
- S6 Video Prompt - What is the Thinking
- S8 Handout 8.0 (Venn diagram)
- S9 Unit 0 Task Summary
- What is a Fraction Original

(2) Session Notes

- Day 1 TIME1 session scripts
- Day 1 TIME 1 Sessions
- Day 2 MS Research
- Day 2 Powerpoint MS
- Day 2 TIME 1 Session Scripts
- Day 2 TIME 1 Sessions

(3) Unit 1 Draft Lessons

- DK Unit 1 - 1A
- DK Unit 1 - 1B

(4) Session Summary Sheet

TIME 2 - December 2013

- KS3 MR lesson study pro-forma v0.3
- Lesson 1a Parts of a shape
- Lesson 1a Slides
- Lesson 1b Pieces of a cake
- Lesson 1b Slides
- Lesson 1c Fair shares
- Lesson 1c Slides

- Lesson 1d Our survey
- Lesson 1d Slides
- Lesson 1e Cards
- Lesson 1e Ordering and Equivalence
- Lesson 1e Slides
- Lesson 1f Cards
- Lesson 1f Slides
- Lesson 1f Milkshakes
- Time 2 Dec 2013 Final Slides
- Time 2 Dec 2013 scripts final
- Time 2 team workshop supporting notes
- Unit 0 Core 1 Feedback document

TIME 3 - February 2014

(1) Lesson 2ef Workshop Presentation Slides

- 2ef TIME Slides Commentary
- Read me first - Suggestion supporting presentation of Lesson 2e&2f
- TIME Presentation Slides 2ef
-

(2) Unit 2 Final Lessons

- **Lesson presentation slides**
    - Lesson 2ef Slides
    - Presentation slides lesson 2b Percentages on the bar model
    - Presentation Slides Lesson 2c Final Identifying Proportional Scenarios
    - Presentation Slides Lesson 2d Final Directly or Inversely Proportional
    - Presentation Slides Lessons 2a Final Contexts To Bar Model

- Lesson 2c Identifying proportional scenarios
- Lesson 2d Directly or inversely proportional
- Lesson 2e
- Lesson 2f
- Unit 2 Lesson 2a
- Unit 2 Lesson 2b

(3) Key project messages for teachers at TIME workshops

(4) Post-Unit Reflection

(5) Suggested TIME workshops presentations for lessons 2ab & 2cd

(6) TIME 3 February 2014 revised Session scripts

(7) TIME 3 February 2014 session slides

TIME 4 -March - April 2014

(1) Handouts for TIME 4 Workshop

- March April TIME workshop Teacher notes and reflection sheet
- March April TIME Teacher summary feedback form
- Post-Unit Reflection
- Unit 0 review Question workshop booklet
- Unit 3 Lesson common issues form
-

(2) Lesson 3ab

- Lesson 3ab Final version
- Lesson 3ab worksheet 1 Recipes
- Lesson 3ab worksheet 2 Lifestyle surveys
- Presentation 3ab 'The ratio table' Stages 1 to 4
- Presentation 3ab 'The ratio table' Stages 5 to 8
-

(3) Lesson 3cd

- Lesson 3cd 9 March 2014
- Lesson 3cd slides 9 March 2014
- UNIT 3cd brie scaling slides 9 March 2014
-

(4) Lesson 3ef

- Lesson3ef Presentation Slides
- Lesson 3ef
- Lesson 3ef Slides
-

(5) TIME 4 March April 2014 session slides

(6) TIME 4 March - April 2014 Session scripts

(7) TIME 5 - June - July 2014

- Multiplicative Reasoning GCSE Q Booklet
- Effective PD Activities Cards
- Session 2 post-unit reflection form
- Session 5 A3 y prop x statements
- Effective aspects of PD sort grid
- Session 6 A3 multiple representations of Q6
- Extension activities hand-out PM sessions
- Teacher project

ANNEX C: Report by Andrew Izsák

Evaluation of the Multiplicative Reasoning Professional Development Project of the National Centre for Excellence in the Teaching of Mathematics

My background: I am a mathematics education researcher who has studied the multiplicative reasoning of practicing and future middle grades mathematics teachers in the United States for over 10 years. I have conducted studies of practicing teachers enacting reform-oriented mathematics curriculum materials in classrooms with their students; studies of practicing teachers working together to deepen their understanding of multiplicative topics in extended professional development courses; studies of future teachers learning to think about multiplication and division, arithmetic with fractions, and ratios and proportional relationships in terms of quantities in specialized content courses; and studies of multiplicative reasoning in a large, national samples of practicing middle grades teachers. My work has been supported through a series of multi-year grants from the National Science Foundation in the United States and has been published in leading mathematics education and applied measurement journals. I also teach a course on multiplicative reasoning to future teachers on a regular basis.

My charge: I was asked to address the following three questions in my evaluation.

1. What is the quality of the professional development activities and materials?

2. How well are the student tests aligned with the content of the professional development?

3. What topics addressed in the professional development should be prioritized for students age 11–14?

1. What is the quality of the professional development activities and materials?

The professional development design had many strengths. It was based on working directly with lead teachers in the TIME team around model, reform-oriented lessons created by three development teams, and having these lead teachers then conduct lesson study with further teachers at local school sites. The lesson study focused on enacting the model lessons and on students' thinking during the lessons. In addition, teachers gained experience interviewing students, providing them a chance to study students' thinking in ways that are not possible by analyzing students' written work alone, or by responding to students' thinking expressed during whole-class lessons. Thus, the professional development placed a heavy, and appropriate, emphasis on teachers learning about opportunities and challenges their students experienced during the model lessons. The model lessons were research-based and emphasized reasoning about multiplicative reasoning in terms of quantities and with various representations of such quantities, not simply computing answers using numerical methods. This approach to improving mathematics instruction is also being pursued intensively in the United States through the development of curriculum materials and new standards for student competencies. Finally, the professional development was focused on critical mathematics taught to students ages 11-14: Multiplication and division, fractions, ratios and proportions lay an essential foundation for the study of subsequent topics in algebra, geometry, statistics, and other areas.

The professional development for the TIME teams was consistent with current understandings of best practices for teacher professional development (e.g., Desimone, 2009; Elmore, 2002; Hill, 2004) as well as research on the effectiveness of professional development (e.g., Desimone, Porter, Garet, Yoon, & Birman, 2002). Desimone (2009) noted general consensus around several characteristics critical for effective professional development. These characteristics include: (a) a focus on subject matter, (b) opportunities for teachers to engage in active learning—for instance, through active discussions and reviewing student work, (c) coherence with teachers' beliefs and external factors such as policies, (d) duration (roughly a minimum of 20 hours, though others recommend a minimum of 40 hours), and (e) collective participation of teachers from the same school or grade level. The materials I reviewed made clear that the professional development addressed four of these five areas directly. The activities I reviewed were squarely focused on mathematical content related to multiplicative reasoning and were well-informed by relevant mathematics education research on children's reasoning about fractions, ratios, and proportions. The lesson commentaries even included summaries of research that informed the designed of the lessons so that teachers could see that the lessons were based on research. The activities themselves made extensive use of drawn models to support a variety of solutions and included suggestions for differentiating instruction for learners who needed different challenges. The participating teachers had opportunities to work on the mathematical tasks together and to consider student responses to those tasks. Furthermore, by engaging in lesson study around the model lessons at their schools teachers had opportunities to work with their colleagues for extended periods of time.

The one area I did not see addressed as directly was teachers' beliefs about the teaching and learning of mathematics. Teachers' beliefs might have played an important role in how they enacted the model lessons with students, but I am not in a position to make judgments about this aspect of the professional development one way or the other. At the same time, the consistent focus in the professional development on using drawn models and on attending to students' thinking provided teachers significant opportunity to reexamine their beliefs about teaching and learning mathematics. For some teachers used to demonstrating algorithms that students then practice, the shift to reform-oriented instructional methods, and beliefs that could support such methods, would likely be difficult and take time. I have seen examples where teachers need to work with reform-oriented materials for a year or more before being convinced of benefits for students.

The professional development for the lead teachers consisted of four meetings spread over one school year. The TIME 1 meeting introduced the lead teachers to the project and its goals. Key mathematical ideas addressed in the TIME 1 meeting included meanings for multiplication and various tools, such as double number lines and ratio tables, that would be used to represent multiplicative relationships in the model lessons. The first model lesson, to be used with students in classrooms, focused on the part-whole fraction construct and partitioning circles and number lines in ways that support comparing, adding, and subtracting fractions. The tasks were clever and posed appropriate challenges for students with a range of prior understandings of arithmetic with fractions.

One aspect of the TIME 1 session was not clear to me was the definition of multiplicative reasoning upon which the project was based. Although an explicit definition was not emphasized in the materials I reviewed, the approach taken was certainly consistent with the notion of the multiplicative conceptual field (Vergnuad, 1983, 1988) that is used widely in mathematics education research. According to Vergnaud, the multiplicative conceptual field consists of a set of interrelated topics that include whole number multiplication and division, fractions, ratios and proportions, linear functions and more. Multiplicative reasoning could then be taken to be reasoning about this field.

Multiplicative relationships are subtle and complicated, and researchers do take different positions on how to characterize and organize them. The materials I reviewed identified two types of multiplicative relationships, repeated addition and scaling. I adopt a slightly different point of view: For me the essential characteristic that distinguishes multiplicative reasoning from additive reasoning is maintaining attention to two different units in a situation, which correspond to the number of groups where all the groups are of the same time. As an example, one can think of 4 recipes where each recipe requires 1/8 lb of butter. One unit is 1/8 lb butter per recipe, and this is a rate, and the other unit is one recipe. In contrast, repeated addition can be interpreted as reasoning with just one unit, the units in each group. In the recipe example, this unit would be 1/8 lb of butter. Maintaining the distinction between number and size of groups can be hard for students and, in my experience, for some teachers. This distinction leads to the two forms of division, measurement and quotitive, that the materials did address explicitly, and supports reasoning about ratios and proportional relationships. The materials I reviewed were not inconsistent with my point of view, but they did not explicitly emphasize everything I think of as essential about multiplicative relationships. It is of course possible that issues I raise were discussed during the meetings.

Another aspect I was not clear about was the use of the term *misconceptions* to characterize some features of student thinking and language about *confronting* misconceptions. Personally, I am persuaded by the arguments of Smith, diSessa, and Roschelle (1993) that efforts to suppress students' ideas runs counter to constructivist approaches to instruction that account for and build upon students' current understandings.

The TIME 2 meeting introduced lead teachers to lesson study and three extended lessons on fractions to be used with teachers as part of lesson study in local school sites. Key tasks for teachers were to anticipate student responses and identify student misconceptions.

Each of the three lessons was developed by a different team, and each had a distinct mathematical emphasis.

The first lesson (lesson ab) was the model lesson introduced at the end of the TIME 1 meeting. Again, main ideas emphasized in this lesson were the part-whole fraction construct and partitioning circles and number lines in ways that support comparing, adding, and subtracting fractions. My one comment about this lesson is that it was not clear the extent to which students would be given explicit support to use what they know about whole-number multiplication and division as a resource for partitioning. In my experience, teachers also need support realizing that whole-number multiplication is a critical resource for becoming fully flexible in partitioning, and not being restricted to partitioning in half, for example.

The second lesson (lesson cd) was explicitly developed within the Realistic Mathematics Education (RME) tradition, a well-established and influential line of Dutch research. In the United States, we have commercially published curricular materials developed within this tradition. The lesson was situated in a sandwich sharing situation that students should be able to imagine. Main ideas emphasized in the second lesson were using drawn models to construct fair shares, attending carefully to the referent unit for various fractions in the situation, and examining the equivalence of different combinations of quantities (e.g., 1/2 + 1/10 of a sandwich and 1/5 + 1/5 + 1/5 of a sandwich). Drawings of sandwiches gradually morphed to bar or strip models, that in turn would morph into ratio tables in the subsequent proportional relationship lesson and, in theory, could also morph into number line models. Students were to make sense of the numeric method of finding common denominators in terms of the sandwich situation and partitioning bars. The tasks were designed to address a common error students make: adding numerators and adding denominators when adding fractions.

The third lesson (lesson ef) was based on two card sorting tasks. The first part of the lesson placed greater emphasis on moving between numerical, linear, and area representations of fractions, percents, and decimals than the first two lessons. The lesson was explicitly designed to address difficulties students have making connections across these different forms of representation and ordering a list of fractions or decimals appropriately. The second part of the lesson emphasized making comparisons between different ratios of juice and milk. One main issue addressed was the difference between forming additive comparisons (through subtraction) and multiplicative comparisons (through division). As one note, I was not clear what point was being made with Charlie: He is correct that the fractions 3/5 and 5/7 are different, which implies that the unit rates and thus the taste of the two milkshakes is different. Perhaps I missed the intended point. The focus on proportional relationships provided a direct lead into the TIME 3 meeting.

The TIME 3 meeting allowed lead teachers to report on results of enacting the three lessons on fractions introduced during the TIME 2 meeting and to study the lessons on ratios and proportions to be used with teachers as part of the next phase of lesson study in local school sites. After reading the materials, I had two questions. First, I wondered whether any of the development teams made explicit use of research by Lamon (e.g., 1994, 1995, 2007) and by Lobato and Ellis (2010) on students' composed unit reasoning. This research describes how students move from focusing on just one of two co-varying quantities to coordinating changes in both quantities appropriately. Maybe the lesson developers were confident that students in this study would already be able to focus on simultaneous changes in two quantities. Discussions of rated addition and moving up and down strips and double number lines seemed consistent with composed unit reasoning, because all of these can be understood as building up new pairs of quantities from an initial batch. Second, I wondered whether the professional development design allowed enough time for teachers to develop a better understanding of what is and what is not a proportional relationship. An extensive research literature exists on students' reasoning about proportional relationships, but there is only a very small literature on teachers' reasoning about proportional relationships (e.g., Orrill & Brown, 2012). The literature on teachers suggest that many face the same challenges as students. Furthermore, in my experience, teachers need extended experiences to develop greater facility reasoning about two quantities that co-vary in a fixed ratio relationship.

As was the case for the TIME 2 meeting, each of the three lessons was developed by a different team, and each had a distinct mathematical emphasis. Consistent with reform-oriented instruction, all three lessons placed greater emphasis on the mathematical structure of the problems, made visible through the use of various drawn models, than on computing answers.

The first lesson (ab) was explicitly developed within the Realistic Mathematics Education (RME) tradition. It extended the use of strip diagrams to solve proportions involving prices of ribbons and other rates; dividing a whole quantity, such as a pizza, into two parts that are in a given ratio; and solving problems about percent increase and decrease.

The second lesson (cd) concentrated on distinguishing between proportional relationships and several other relationships that students and teachers have confused with proportional relationships in the research literature. These included affine, constant value, and indirectly proportional relationships. The tasks were designed to address common student errors, like thinking that two quantities are in a proportional relationship if they are both increasing, confusing directly proportional relationships with other relationships, and using additive rather than multiplicative comparisons, inappropriately.

The third lesson (ef) examined the difference between proportional relationships and several other relationships that students and teachers have confused with proportional relationships in the research literature. These included affine, constant difference, indirectly proportional, and quadratic relationships. To help highlight the difference, the lesson included both tables of values and a double number line that I have seen referred to as the parallel axis representation for functions. This emphasis on connections across difference representations was consistent with the emphasis of the third lesson on fractions discussed above.

The TIME 4 meeting allowed lead teachers to report on results of enacting the three lessons on ratios and proportions introduced during the TIME 3 meeting and to study the lessons on wider applications of multiplicative reasoning to be used with teachers as part of the final phase of lesson study in local school sites.

The first lesson (ab) was explicitly developed within the Realistic Mathematics Education (RME) tradition. This lesson replaced the bar or strip model with the ratio table, which can be thought of as a bar or strip without information about the relative size of the recorded quantities. Thus ratio tables can be more convenient to work with depending on the range of numbers being recorded. The lesson illustrated how a range of topics can be unified through use of the ratio table. Topics included comparing rates, measurement and partitive division, and percentages.

The second lesson (cd) used double number lines to help students see the same division structure across problems in which the numbers vary. Thus, students can think about partitive and quotitive division when the divisor is greater than the numerator. The second part of the lesson addressed multiplication interpreted in three ways: repeated addition, arrays, or scaling. A main goal was to address students' common generalization that multiplication makes larges and division makes smaller.

The third lesson (ef) presented a situation about an orange grove with multiple quantities and relationships among them. A main goal of the lesson was for students to practice identifying appropriate arithmetic operations as the numbers change—for instance, from whole numbers to decimals greater than one or less than one.

To summarize, the professional development appears to have been well-aligned with our current understandings of best practices and targeted critical mathematics content. The activities for students emphasized using drawn models to reason about quantities in situations, not on numerical methods for computing answers. It is possible that for some teachers, transitioning to such materials is likely to be challenging and to take time, perhaps even more time than the current project allowed. Researchers take various positions about how to characterize and organize multiplicative reasoning and learning. My own views and those of the lesson developers appear well-aligned in most but not all cases.

2. How well are the student tests aligned with the content of the professional development?

I assume I am being asked this question because one would like to know the extent to which effects of the professional development on students' multiplicative reasoning could be detected using student test data. There are at least two difficulties here. First, establishing empirical connections between teachers' mathematical knowledge, which influences how they enact lessons, and students' achievement has been very difficult historically, though there are at least two main examples of recent successes (Baumert et al., 2010; Hill, Rowan, & Ball, 2005). Second, judging alignment leads directly to a central tension in research on transfer: Are connections between the model lessons and test items viewed from a student's point of view or an expert's point of view? If students have a good understanding of ratios and proportional relationships in the model lessons that use strip diagram, but do not see strip diagrams in the test questions, they might not activate ways of reasoning developed in the model lessons and thus answer items incorrectly. In this case, incorrect responses to test items would underrepresent what students had learned.

My overall judgment is that the *Progress in Maths* tests are distal measures of the content in the model lessons. Thus, I think it possible that these tests would be rather insensitive to gains that students might have made as they participated in the model lessons. I examined all items in the three tests I was sent and placed each item into one of four categories. The *Not Multiplicative Reasoning* (Not MR) category consists of those items I judged not to highlight multiplicative relationships. The *Not Related Multiplicative Reasoning* (Not Related MR) category consists of items I judged could be interpreted in terms of repeated groups, but that did not connect directly to topics covered in the model lessons. The *Weakly Related Multiplicative Reasoning* (Weak MR) category consists of those items that addressed a topic also addressed in the model lessons, but without some of the same key features such as specific drawn models. Finally, the Strongly Related Multiplicative Reasoning (Strong MR) category consists of those items that addressed a topic also addressed in the model lessons *and* that included similar representations. Note that the distinction between weakly and strongly items is informed by the discussion of transfer above. I went through the test items twice to check my categorization, but ideally one would like at least one more rater to assess the reliability. It is certainly possible that I have overlooked some connections between the model lessons and the test items. As shown in Tables 1–3, there were just a handful of items I judged to be strongly related to the content of the model lessons.

The three tables below summarize my categorization by test. Short phrases in the cells indicate my reasons for each judgment. The following notes explain some of my decisions in more detail.

1. Items that required attention to factors and multiples, but only numerically, I put in the not related category because an important feature of the model lessons was for students to use problem situations to explain solution methods, even they already new how to compute a particular answer using some recalled algorithm.
2. Because one set of model lessons highlighted connections across multiple representations, I put test items with a similar emphasis in the weakly related group, even if they did not address multiplicative relationships.
3. By reciprocal relations, I mean understanding that if A is 4/7 of B, then B is 7/4 of A. Being able to form such relationships is critical to reasoning multiplicatively about measurement. I saw an example of reciprocal relations in the coloured rods task in the TIME 1 materials, but I did not see this as a focus of the model lessons students would have experienced. The closest point of contact were questions about a percent increase followed by a percent decrease or vice versa.

Table 1. Alignment of Items in Progress in Maths 12

Item	Not MR	Not Related MR	Weak MR	Strong MR
1a	Subtraction of whole #s			
1b		Calculate quotient of whole numbers		
2	nets			
3a		Calculate average		
3b		Calculate		

		average		
4		Whole number factors		
5a	Subtraction			
5b			Build-up method	
6	Read graph			
7		Calculate quotient of whole numbers		
8		Finding perimeter from area; Division		
9a		Multiply to find area (10 * 16)		
9b		Multiply to find area (4 * 16)		
9c	Satisfy two constraints			
10a	Read graph			
10b			Find percent	
11	Counting to predict path in plane.			
12	Balance and equations			
13a	Read table			
13b			Connect representations of percent	
14		Writing expressions that use multiplication		
15		Computation		
16	Visualization of shapes			
17		Computation		
18	Classifying shapes			
19a		Reciprocal relations		
19b	Examining alternative cases			
20	Constant difference			
21			Numerical subtraction of fractions	
22	Visualization of shapes			
23	Subtraction of whole #s			
24			Connect	

			representations of weight	
25			Multiplication as model of repeated groups	
26		Reciprocal relations		
27			Measurement division	
28		Algebra and averages		

In summary, of the *28 Progress in Maths 12* test questions, 7 were weakly related and none were strongly related to the model lessons.

Table 2. Alignment of Items in Progress in Maths 13

Item	Not MR	Not Related MR	Weak MR	Strong MR
1			Fraction multiplication	
2			Locating on number line	
3		Calculate average		
4		Scale drawing; area		
5		Quotient of whole numbers		
6	Read graph			
7				Scaling recipe by 3/2.
8	Subtraction			
9			Multiplication and division of two quantities in fixed ratio	
10	Geometric properties			
11	Read table			
12	Number properties			
13	nets			
14	Subtraction			
15				Part whole relationships using area
16a	Writing additive expression			
16b		Writing multiplicative expression		
17	Read graph			
18			Numerical subtraction of fractions	
19		Combinations		
20			Missing value	
21		Calculate quotient of whole numbers		
22		Lengths and areas of rectangles		
23		Probability		
24a		Compute average		
24b			Translating across	

			representations	
25				Reciprocal relationships of percents
26	Visualizing shapes			
27a		Repeated addition;		
27b		Reciprocal relations		
28		Volume		
29				Representing percents on pie chart
30a	Write expression for addition			
30b		Write expression for multiplication and subtraction		

In summary, of the 30 *Progress in Maths 13* test questions, 6 were weakly related and 4 were strongly related to the model lessons.

Table 3. Alignment of Items in Progress in Maths 14

Item	Not MR	Not Related MR	Weak MR	Strong MR
1a	Addition			
1b			Find percent	
2		Factors; reciprocal relations		
3		Estimation and scale drawings**?**		
4a		Compute average		
4b				Percent change
5		Repeated addition in sequence		
6			Find percent	
7a		Compute quotient		
		Compute product		
8a			Find percent	
8b			Find percent	
8c		Probability		
9	Recall π			
10				Missing value; Unit conversions and rates
11	Geometry			
12	Subtraction			
13		Reciprocal relations		
14	Nets			
15		Compare fractions		
16				Measurement division using lengths
17	Visualizing shapes			
18		Calculate quotient of whole numbers		
19	Read graphs			
20	Write expressions for addition			
21a		Multiply to find area (15 * 4)		
21b	Satisfying constraints			
22		Reciprocal relations		
23		Algebra and averages		
24	Properties of			

	polygons			
25		Solving equations in geometry context		
26		Writing expressions for multiplication		
27			Build-up method	
28		Interpreting computations using multiplication		
29		Writing expressions for multiplication		

In summary, of the 29 *Progress in Maths 14* test questions, 5 were weakly related and 3 were strongly related to the model lessons. The relatively low counts of items weakly or strongly related to the model lessons is the primary reason that I judge the *Progress in Maths* tests to be distal measures of what students might have learned during the model lessons. Furthermore, the model lessons approached more conceptual aspects of multiplicative relationships than did the student tests: for instance asking questions about the appropriate referent unit for a number, comparing ratios of two mixtures, distinguishing proportional relationships from other types of relations, and attending to the meaning of place value when comparing decimals.

3. What topics should be prioritized for students age 11-14?

This is the most difficult of the three questions. Researchers have proposed various entry points into fractions and related multiplicative topics, and I doubt there is one that minimizes potential difficulties students experience while at the same time maximizes the accessibility of fractions, ratios, proportional relationships, etc. The model lessons I reviewed were not intended as a complete curriculum, but rather as illustrative examples of lessons that could support reform-oriented instruction and teachers' access to their students' thinking. Generally speaking, one might expect lessons on fractions to preceded lessons on ratios and proportional relationships, but one could argue that ratios and proportional relationships that do not require fractions could be introduced first. Having said that, I did identify two themes that could be used to connect the lessons I reviewed in learning trajectories that span more than one grade. These themes are not a direct answer to the question I was asked, but they do provide a way to think about how to organize the model lessons in a more fully developed curriculum that spans multiple grades.

The first theme has to do with recognizing and appropriately distinguishing arithmetic operations across situations. It is well-known that students do not "conserve" operations when numbers change. When I teach future teachers, I address this issue using a combination of three things. First, I introduce an explicit definition of multiplication that highlights the distinction between the number of groups (multiplier) and the size of the groups (multiplicand). Second, I use this definition of multiplication as the criterion against which to judge whether a situation calls for additive or multiplicative comparisons and for identifying whether the situation calls for multiplication or division. Comparisons based on multiple groups of the same size can me made multiplicatively. Third, I use number line and area representations to help future teachers see connections across problems with different numbers. As one example, when discussing division, I emphasize that questions asking *how many are in one group* determine partitive division and use double number lines to show that the connection between partitive division when the divisor is greater than and less than one. One can start anywhere on the double number line and find out how many units of quantity A go with one unit of quantity B. Whether you start with 4 units or with 3/5 units of quantity B does not change the question and does not change the arithmetic operation. I find that future teachers can not only make sense of this treatment but also readioy extend reasoning with double numbers to solve proportions in general. Thus, partitive division with various sized divisors and solving proportions are integrated, where double number lines support that integration.

The second theme has to do with the use of representations in instruction. Of the materials I reviewed, I was especially struck by those developed within the RME tradition. These materials reflected close attention to ways that forms of representation (bar or strip diagrams) can be integral to students' reasoning about a particular topic, and they paid close attention to the evolution of strip diagrams into ratio tables. One could also imagine the evolution of strip diagrams into number lines. I do think it is important to consider how curriculum materials support larger trajectories for representations. An especially nice example I have seen is that of Japanese curriculum materials translated into English. These materials have a trajectory for length-based representations that lead to number lines over several grades. A main advantage I see in this approach is that students have ample time to learn how to think with particular forms of representation, so that they can transition from learning how to interpret representations to using representations as tools for thinking about new situations. This is akin the model of vs. model for distinction made within the RME tradition. At the same time, I also see value in having students make connections across multiple representations, as emphasized in some of the other model lessons, but I think the field has yet to address important questions. As one example, at least in the United States, it is not uncommon to hear teachers explain that the purpose of using multiple representations or solution methods in a lesson is for students with different learning styles to find one representation or method that they can use successfully. In this case, multiple representations in a classroom can become just one representation per child. I certainly agree that different children will find different representations and methods more accessible, but I also think that building connections across representations deepens understanding because students can see how they same idea can be projected in different ways. In my view, the field still has much to learn about ways to use representations effectively in instruction, including how to balance developing particular representations into tools for thinking with developing connections across multiple representation and seeing different projections of core concepts.

References:

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*, 181–199. doi:10.3102/0013189X08331140

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis, 24*, 81–112. doi:10.3102/01623737024002081

Elmore, R. F. (2002). *Bridging the gap between standards and achievement: The imperative for professional development in education.* Washington, DC: Albert Shanker Institute.

Hill, H. C. (2004). Professional development standards and practices in elementary school mathematics. *Elementary School Journal, 104*, 215–231.

Hill, H., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371-406.

Lamon, S.J. (1994). Ratio and proportion: Cognitive foundations in unitizing an norming. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 89-120). Albany, NY: State University of New York Press.

Lamon, S.J. (1995). Ratio and proportion: Elementary didactical phenomenology. In J. T. Sowder & B. P. Schappelle (Eds.), *Providing a foundation for teaching mathematics in the middle grades* (pp. 167-198). Albany, NY: State University of New York Press.

Lamon, S. (2007). Rational numbers and proportional reasoning: Toward a theoretical framework for research. In K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 629–667). Charlotte, NC: Information Age Publishing.

Lobato, J., & Ellis, A. B. (2010). *Essential understandings: Ratios, proportions, and proportional reasoning.* In R. M. Zbiek (Series Ed.), *Essential understandings*. Reston, VA: National Council of Teachers of Mathematics.

Orrill, C. H., & Brown, R. E. (2012). Making sense of double number lines in professional development: Exploring teachers' understandings of proportional relationships. *Journal of Mathematics Teacher Education, 15*(5), 381–403.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993/1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115-163.

Vergnaud, G. (1983). Multiplicative structures. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and processes* (pp. 127–174). New York: Academic Press.

Vergnaud, G. (1988). Multiplicative Structures. In J. Hiebert & M. Behr (Eds.), *Number concepts and operations in middle grades* (pp. 141–161). Reston, VA: National Council of Teachers of Mathematics; Hillsdale, NJ: Erlbaum.

Annex D: Technical Appendix

Contents

D1: Reading Multilevel Model Tables & Calculating Hedges g effect size statistics.

To assist with the reading of the statistical findings for the impact analyses multilevel models which are summarised within Sections 7.3 to 7.8 and detailed in Appendices E2 to XX, this section of the technical Appendix provides a glossary of terms used within the models and a guide on how the hedges g effect size statistics were calculated.

Part A: Glossary for reading multilevel model tables.

For each reported model, two tables are provided. The first of these summarises the model itself and presents the effect size statistics. Following this, the variance structures are displayed. This glossary takes each table in turn to try to assist readers to interpret the statistical analyses conduced within the impact analyses.

To help illustrate points within this glossary and later on with the calculation of effect size statistics, the first model shown in Appendix E2 (and summarised in section 7.3 in the main report) is used - this is the model looking at the combined PiM 12 to 14 outcome variable.

1. The Model Table

Coef: This is the estimated coefficient term for each of the explanatory variables. This term can be positive, zero or negative.

A positive coefficient indicates a higher score on the outcome variable. For example, a positive coefficient of '+0.5' is seen for the gender variable in the final model shown in Table 1 of Appendix E2. This tells us that when taking other explanatory variables included in the model into account, on average females scored higher than males on the outcome variable by 0.5 points.

Similarly, a negative coefficient indicates a lower score on the outcome variable compared with their male peers. For example, in Appendix E2, Table 1, a negative coefficient of '-1.4' is seen for the FSM variable, this tells us that when taking other explanatory variables included in the model into account, on average FSM participants attained lower scores on the outcome variable compared with their non-FSM peers by 1.4 points.

A zero coefficient indicates that there is no differences in the outcome variable with respect to an explanatory variable. For example, if a zero coefficient is seen for the pre-test variable, this tells us that scores on the outcome variable are not statistically associated with pre-test scores - although for the MRP analyses, this is not the case as a positive coefficient of +11.3 points is observed (A difference of 1 point in KS2 maths attainment is associated with a difference of 11.3 points in the overall PiM 12 to 14 outcome measure).

Coefficients are shown in the units of the outcome variable. This means that comparing coefficients of one explanatory variable across multiple models is problematic (because of the differing units). This is one reason why effect size statistics are used to help interpret models. These standardise the units and enable direct comparisons to be made (as is shown below).

s.e. This is the standard error for each coefficient term. These are used to give an indication on how precise the estimated coefficient term is (smaller values indicate greater precision). The standard error takes account of sample size and are widely used in tests of statistical significance and calculation of confidence intervals. One method of checking whether a coefficient is statistically significant is to compare the coefficient value with its associated standard error - if the coefficient value is more than 1.96 (or roughly twice) the size of the standard error, the coefficient will be statistically significant.

Constant This is the average (mean) of the outcome variable once all of the explanatory variables have been included. This represents the mean score for the reference group of the model - i.e. non-FSM males who attained a mean score on their KS2 maths test. To calculate the mean for a different group, a coefficient term would need to be added - e.g. for non-FSM females who attained a mean KS2 maths attainment score, the gender coefficient would need to be added.

Effect Size: This is a standardised measure that is calculated from the model coefficient. Hedges g is the effect size measure used (see below for more detail on calculating this). The measure standardises so that units are converted into standard deviations - which, unlike the raw coefficient, can be directly compared across many models. Higher values indicate greater statistical impact.

2. Variance Structure and Explanatory Power

The second table provided for each model summarises the variance structure and explanatory power of the model. The following terms are used within these tables:

-2 Log Likelihood This is the total amount of variation in the outcome variable. Two values are provided; the empty model and parsimonious model. The log-likelihood for the empty model displays the amount of variation before the model is constructed and the parsimonious model displays the variation remaining after the final model has been fitted.

School level variance: This is the amount of variation located at the school-level. This is also shown for the empty and parsimonious model stages as absolute values and as a proportion of the total variation.

Class level variance: This is the amount of variation located at the class-level. This is also shown for the empty and parsimonious model stages as absolute values and as a proportion of the total variation.

Pupil level variance: This is the amount of variation located at the individual pupil-level. This is also shown for the empty and parsimonious model stages as absolute values and as a proportion of the total variation.

Intra-Cluster Correlation (ICC)- This is the proportion of variation that is found at the higher levels of the model - i.e. the school and class levels.

Explanatory Power: Statistically, the explanatory power of a model is the proportion of variation in an outcome variable that can be accounted for by variations across explanatory variables. For a multilevel model, there are four explanatory power statistics - known as R-square or R^2.

R^2 - This is the overall explanatory power across all levels - i.e. the proportion of variation in the outcome variable that is accounted for by variations across the explanatory variables. In Appendix E2, the overall explanatory power is calculated as 70% - 70% of the total variation in the PiM 12 to 14 combined outcome measure is accounted for through variations across the explanatory variables.

R_{School}^2: This is the proportion of variation in the outcome variable at the school level that is accounted for by variations across the explanatory variables. In Appendix E2, the school level explanatory power is calculated as 93% - 93% of the school level variation in the PiM 12 to 14 combined outcome measure is accounted for through variations across the explanatory variables.

R_{Class}^2: This is the proportion of variation in the outcome variable at the class level that is accounted for by variations across the explanatory variables. In Appendix E2, the class level explanatory power is calculated as 89% - 89% of the class level variation in the PiM 12 to 14 combined outcome measure is accounted for through variations across the explanatory variables.

R_{Pupil}^2: This is the proportion of variation in the outcome variable at the individual pupil level that is accounted for by variations across the explanatory variables. In Appendix E2, the pupil level explanatory power is calculated as 23% - 23% of the pupil level variation in the PiM 12 to 14 combined outcome measure is accounted for through variations across the explanatory variables.

Looking at Appendix E2, Table 1 in the empty model, the distribution of the variation in the overall PiM 12 to 14 measures is observed to be partitioned between the three levels with 18% located at the school level, 52% at the class level and the remaining 30% at the individual level. For the final (parsimonious) model, the remaining variation is observed to be partitioned such that 4% is located at the school level, 19% is at the class level and the remaining 77% at the individual level. This is echoed in the explanatory power statistics that tell us that the vast majority of school level (93%) and class level (89%) variation is accounted for in the model whilst at the pupil level, explanatory power is weaker (23%). So whilst around 70% of the variation in the PiM 12 to 14 overall outcome variable is located at the class and school levels combined, once the model is fitted most of this higher level variance is accounted for. This leaves most of the unexplained variance located at the pupil level (i.e. 77%).

Part B: Calculating Effect Size Statistics.

An effect size is a statistical estimate of the strength of a phenomenon in standardised units. In the context of this research, the effect size provides an indication of the difference between the intervention and control groups for the PiM outcome measures. Whilst the model coefficients do also provide an indication of this, the effect size standardises these coefficients so that they can be compared directly with each other and across other research studies. Without standardisation, the size of coefficient is dependent on the scale and units of the outcome measure and so it is not possible to compare these directly.

The effect size used is Hedges g. This is a similar effect size statistic to Cohen's d but uses a standard deviation that is pooled between the intervention and control groups. It also includes a slight correction to reduce the bias associated with Cohen's d when dealing with small samples.

The combined PiM12to13 variable is used in this illustration.

Table 1: Primary Outcome: PiM12to13 Summary Statistics

	Control	Intervention	All Respondents
Mean	103.0	97.9	100.3
Standard deviation	16.15	13.14	14.87
n=	3,138	3,427	6,565
Pooled standard	14.65		

	Control	Intervention	All Respondents
deviation*			

Referring to the parsimonious model for the PiM12to13 combined outcome measure (Appendix E2, Table 1), the model shows a coefficient of -1.0 for the intervention group with a standard error of 0.66.

The standard error can be used to calculate 95% confidence intervals for the coefficient:

95% Confidence Intervals:

Coefficient +/- (1.96 x standard error) = -1.0 +/- (1.96 x 0.66)

Upper limit of confidence interval = - 1.0 + 1.3 = + 0.30

Lower limit of confidence interval = -1.0 - 1.3 = - 2.3

The above coefficient and upper / lower confidence intervals can be converted into an effect size by dividing by the standard deviation. For Cohens d, this would be the standard deviation of the primary outcome for all respondents (i.e. s=14.87) whilst for Hedges g, the pooled standard deviation is used (s=14.65).

The pooled standard deviation is calculated using the following formula:

Converting the coefficient into the (Hedge's g) effect size:

Hedges g	= -1.0 / 14.65	= - 0.07
Hedges g confidence interval upper limit	= + 0.3 / 14.65	= + 0.02
Hedges g confidence interval lower limit	= - 2.3 / 14.65	= - 0.15

In summary, an effect size of (h=) -0.07 with 95% confidence intervals between -0.15 and +0.02.

Since these confidence intervals cross the zero bound - the coefficient is not statistically significant. This means that once sample size and random variation has been taken into account, the effect size is not sufficiently large to conclude a genuine impact.

D2: Maths Attainment (GL PiM 12 to 14 Standardised Age Score) Models (Intention to Treat Analyses).

(Summarised in Section 7.3 in the main report)

NOTE - to assist in reading these tables, please refer to Appendix E1 (glossary & effect sizes)

Table 1a: **Primary outcome = GL PiM12to13 Combined Measure (all areas combined)**

Intention to Treat Analyses

n=6,565 in 349 classes in 58 secondary schools

	Empty (Null) Model		**Control / Intervention Alone**		**Pupil & School Level Variables**		**Final Model**	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	**-4.1**	**2.07**	-1.0	0.67	-1.0	0.66
Pre-test (KS2 Maths)	-	-	-	-	**11.4**	**0.20**	**11.3**	**0.20**
FSM (eligible & claiming)	-	-	-	-	**-1.4**	**0.24**	**-1.4**	**0.24**
Gender (female)	-	-	-	-	**0.5**	**0.20**	**0.5**	**0.20**
SEN (has SEN)	-	-	-	-	**-1.9**	**0.30**	**-1.9**	**0.30**
School Level Attainment	-	-	-	-	**-2.9**	**1.11**	**-2.9**	**1.11**
School level (selective	-	-	-	-	**10.1**	**1.55**	**10.1**	**1.55**
School level (%FSM)	-	-	-	-	**-0.1**	**0.04**	**-0.1**	**0.03**
School level (IDACI)	-	-	-	-	1.0	3.33	/	/
School level (OFSTED)	-	-	-	-	**2.9**	**1.03**	**2.9**	**1.03**
Constant	**97.4**	**1.07**	**99.5**	**1.48**	**100.7**	**1.03**	**100.7**	**1.02**

Effect sizes

Hedges g	-	**-0.28***	-0.07	-0.07
95% CIs		**(-0.56, -0.01)**	(-0.16, +0.02)	(-0.15, +0.02)

Table 1b: ***Variance structure & explanatory power*** **- GL PiM12to13 Combined Measure Intention to Treat Analysis (Overall Maths Attainment)**

All Regions / Year Groups combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	-23,930.6		-22,781.4		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level Variance	42.5	18%	2.9	4%	93%
Class Level Variance	121.4	52%	13.7	19%	89%
Pupil Level Variance	71.3	30%	55.1	77%	23%
Total (all levels combined)	235.1	100%	71.6	100%	70%

Table 2a: Primary outcome = GL PiM12 Y7 (all areas combined)

Intention to Treat Analyses

n=2,388 in 128 Y7 classes in 55 secondary schools

	Empty (Null) Model		**Control / Intervention Alone**		**Pupil & School Level Variables**		**Final Model**	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	-2.0	2.74	-0.4	0.72	-0.3	0.76
Pre-test (KS2 Maths)	-	-	-	-	**12.9**	**0.28**	**12.8**	**0.28**
FSM (eligible & claiming)	-	-	-	-	**-2.2**	**0.40**	**-2.2**	**0.40**
Gender (female)	-	-	-	-	**0.9**	**0.33**	**0.9**	**0.33**
SEN (has SEN)	-	-	-	-	**-2.1**	**0.49**	**-2.2**	**0.49**
School Level Attainment	-	-	-	-	**-2.7**	**1.21**	/	/
School level (selective	-	-	-	-	**6.2**	**1.63**	**6.3**	**1.71**
School level (%FSM)	-	-	-	-	-0.1	0.04	**-0.1**	**0.03**
School level (IDACI)	-	-	-	-	-1.0	3.50	/	/
School level (OFSTED)	-	-	-	-	**2.0**	**1.11**	/	/
Constant	**98.5**	**1.38**	**99.5**	**1.96**	**102.7**	**1.14**	**101.5**	**0.77**
Effect sizes								
Hedges g	-		-0.13		-0.03		-0.02	
95% CIs			(-0.49, +0.23)		(-0.12, +0.07)		(-0.12, +0.08)	

Table 2b: *Variance structure & explanatory power* **- GL PiM12 Y7**

Intention to Treat Analysis (Overall Maths Attainment)

Y7 subsample - All regions combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	-8.962.8		-8,271.2		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level	51.6	21%	3.7	6%	93%
Class Level	101.3	42%	3.8	6%	96%
Pupil Level	89.4	37%	56.3	88%	37%
Total (all levels combined)	242.3	100%	63.8	100%	74%

Table 3a: Primary outcome = GL PiM13 Y8 (all areas combined)

Intention to Treat Analyses

n=2,076 in 110 Y8 classes in 56 secondary schools

	Empty (Null) Model		**Control / Intervention Alone**		**Pupil & School Level Variables**		**Final Model**	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	**-6.1**	**2.98**	-1.2	1.17	-1.3	1.17
Pre-test (KS2 Maths)	-	-	-	-	**10.5**	**0.38**	**10.4**	**0.38**
FSM (eligible & claiming)	-	-	-	-	**-1.0**	**0.43**	**-1.0**	**0.43**
Gender (female)	-	-	-	-	0.4	0.35	/	/
SEN (has SEN)	-	-	-	-	**-2.0**	**0.51**	**-2.1**	**0.51**
School Level Attainment	-	-	-	-	**-3.7**	**1.85**	**-3.7**	**1.86**
School level (selective	-	-	-	-	**10.7**	**2.67**	**10.8**	**2.68**
School level (%FSM)	-	-	-	-	-0.1	0.08	**-0.1**	**0.05**
School level (IDACI)	-	-	-	-	1.5	6.13	/	/
School level (OFSTED)	-	-	-	-	3.3	1.69	**3.3**	**1.70**
Constant	**97.8**	**1.56**	**100.9**	**2.07**	**102.0**	**1.79**	**102.2**	**1.79**
Effect sizes								
Hedges g	-		**-0.40***		-0.08		-0.08	
95% CIs			**(-0.79, -0.01)**		(-0.23, +0.07)		(-0.24, +0.07)	

Table 3b: ***Variance structure & explanatory power*** **- GL PiM13 Y8**

Intention to Treat Analysis (Overall Maths Attainment)

Y8 subsample - All regions combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	-7,488.6		-7,205.6		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level	67.5	27%	7.8	10%	88%
Class Level	117.7	47%	14.5	19%	88%
Pupil Level	65.0	26%	54.4	71%	16%
Total (all levels combined)	250.2	100%	76.7	100%	69%

Table 4a: Primary outcome = GL PiM14 Y9 (all areas combined)

Intention to Treat Analyses

n=2,101 in 111 Y9 classes in 52 secondary schools

	Empty (Null) Model		**Control / Intervention Alone**		**Pupil & School Level Variables**		**Final Model**	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	**-4.6**	**2.29**	-1.4	1.18	-1.6	1.24
Pre-test (KS2 Maths)	-	-	-	-	**8.4**	**0.42**	**8.3**	**0.42**
FSM (eligible & claiming)	-	-	-	-	**-0.8**	**0.42**	**-0.9**	**0.41**
Gender (female)	-	-	-	-	-0.1	0.34	/	/
SEN (has SEN)	-	-	-	-	**-1.3**	**0.51**	**-1.3**	**0.51**
School Level Attainment	-	-	-	-	-4.4	2.36	/	/
School level (selective	-	-	-	-	**16.5**	**2.87**	**17.6**	**2.92**
School level (%FSM)	-	-	-	-	-0.1	0.07	/	/
School level (IDACI)	-	-	-	-	3.9	5.71	/	/
School level (OFSTED)	-	-	-	-	4.4	2.21	/	/
Constant	**95.9**	**1.19**	**98.3**	**1.66**	**98.1**	**1.67**	**97.1**	**0.94**
Effect sizes								
Hedges g	-		**-0.34***		-0.10		-0.11	
95% CIs			**(-0.67, -0.01)**		(-0.27, +0.07)		(-0.29, +0.06)	

Table 4b: *Variance structure & explanatory power* **- GL PiM14 Y9**

Intention to Treat Analysis (Overall Maths Attainment)

Y9 subsample - All regions combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	-7,422.6		-7,249.3		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level	4.6	2%	1.1	1%	76%
Class Level	141.1	70%	33.2	39%	76%
Pupil Level	56.2	28%	50.9	60%	9%
Total (all levels combined)	201.9	100%	85.3	100%	58%

D3: PiM 12 to 14 Subscale Models (Intention to Treat Analyses).

(Summarised in Section 7.4 in the main report)

Table 1a: Primary outcome = GL PiM12 Y7 - Subscale 1 (ANY PiM MR ITEMS)

Y7, All regions combined.

Intention to Treat Analyses

n=2,388 in 128 Y7 classes in 55 secondary schools

	Empty (Null) Model		**Control / Intervention Alone**		**All Pupil & School Level Variables**		**Final Model**	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	-0.8	1.30	-0.1	0.38	0.0	0.41
Pre-test (KS2 Maths)	-	-	-	-	**6.0**	**0.14**	**6.0**	**0.14**
FSM (eligible & claiming)	-	-	-	-	**-1.2**	**0.19**	**-1.2**	**0.19**
Gender (female)	-	-	-	-	**0.6**	**0.16**	**0.6**	**0.16**
SEN (has SEN)	-	-	-	-	**-0.9**	**0.24**	**-0.9**	**0.24**
Age (in months)	-	-	-	-	0.02	0.02	/	/
School Level Attainment	-	-	-	-	-1.2	0.64	/	/
School level (selective	-	-	-	-	**2.4**	**0.86**	**2.9**	**0.91**
School level (%FSM)	-	-	-	-	0.0	0.02	/	/
School level (IDACI)	-	-	-	-	-0.4	1.86	/	/
School level (OFSTED)	-	-	-	-	1.2	0.59	/	/
Constant	**12.8**	**0.65**	**13.2**	**0.93**	**12.0**	**3.23**	**13.6**	**0.32**

Effect sizes

Hedges g	-	-0.11	-0.01	0.00
95% CIs		(-0.46, +0.25)	(-0.12, +0.09)	(-0.11, +0.12)

Table 1b: ***Variance structure & explanatory power*** **- GL PiM12 Y7 Subscale 1**

Intention to Treat Analysis

Y7 subsample - All regions combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	-7,186.7		6,535.2		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level	10.9	20%	1.1	7%	90%
Class Level	23.6	43%	1.3	8%	95%
Pupil Level	20.2	37%	13.0	84%	36%
Total (all levels combined)	54.7	100%	15.4	100%	72%

Table 2a: **Primary outcome = GL PiM12 Y7 - Subscale 2**

PiM MR items that are weakly or strongly connected with the MRP

Y7, All regions combined.

Intention to Treat Analyses

n=2,388 in 128 Y7 classes in 55 secondary schools

	Empty (Null) Model		Control / Intervention Alone		Pupil & School Level Variables		Final Model	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	-0.2	0.37	0.0	0.10	0.0	0.10
Pre-test (KS2 Maths)	-	-	-	-	**1.8**	**0.05**	**1.8**	**0.05**
FSM (eligible & claiming)	-	-	-	-	**-0.3**	**0.08**	**-0.3**	**0.08**
Gender (female)	-	-	-	-	0.0	0.06	/	/
SEN (has SEN)	-	-	-	-	**-0.2**	**0.09**	**-0.2**	**0.09**
Age (in months)	-	-	-	-	0.0	0.01	/	/
School Level Attainment	-	-	-	-	-0.2	0.18	/	/
School level (selective	-	-	-	-	**0.7**	**0.23**	**0.7**	**0.24**
School level (%FSM)	-	-	-	-	0.0	0.01	/	/
School level (IDACI)	-	-	-	-	0.0	0.50	/	/
School level (OFSTED)	-	-	-	-	0.1	0.16	/	/
Constant	**3.7**	**0.18**	**3.8**	**0.26**	**4.3**	**1.26**	**4.1**	**0.10**

Effect sizes

Hedges g	-	-0.09	0.00	0.00
95% CIs		(-0.41, +0.24)	(-0.09, +0.09)	(-0.09, +0.09)

Table 2b: *Variance structure & explanatory power* **- GL PiM12 Y7 Subscale 2**

Intention to Treat Analysis

Y7 subsample - All regions combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	-4,709.9		-4,285.1		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level	0.9	18%	0.1	2%	94%
Class Level	1.7	33%	0.1	2%	97%
Pupil Level	2.6	50%	2.1	95%	21%
Total (all levels combined)	5.3	100%	2.2	100%	59%

Table 3a: Primary outcome = GL PiM13 Y8 - Subscale 1 (ANY MR ITEMS)

Y8, All regions combined.

Intention to Treat Analyses

n=2,076 in 110 Y8 classes in 56 secondary schools

	Empty (Null) Model		**Control / Intervention Alone**		**Pupil & School Level Variables**		**Final Model**	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	-3.4	1.88	-0.6	0.87	-0.2	0.92
Pre-test (KS2 Maths)	-	-	-	-	**6.2**	**0.25**	**6.2**	**0.24**
FSM (eligible & claiming)	-	-	-	-	**-0.6**	**0.26**	**-0.6**	**0.26**
Gender (female)	-	-	-	-	0.0	0.22	/	/
SEN (has SEN)	-	-	-	-	**0.7**	**0.32**	**0.7**	**0.31**
Age (in months)	-	-	-	-	-0.1	0.03	/	/
School Level Attainment	-	-	-	-	-2.5	1.37	/	/
School level (selective	-	-	-	-	**7.1**	**1.98**	**8.0**	**2.04**
School level (%FSM)	-	-	-	-	-0.1	0.06	/	/
School level (IDACI)	-	-	-	-	1.0	4.58	/	/
School level (OFSTED)	-	-	-	-	1.8	1.26	/	/
Constant	**13.8**	**0.98**	**15.5**	**1.33**	**24.4**	**4.84**	**14.7**	**0.68**

Effect sizes

Hedges g	-	-0.36	-0.06	-0.02
95% CIs		(-0.74, +0.03)	(-0.24, +0.12)	(-0.21, +0.17)

Table 3b: ***Variance structure & explanatory power*** **- GL PiM13 Y8 Subscale 1**

Intention to Treat Analysis

Y8 subsample - All regions combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	-6,483.9		-6.225.3		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level	25.9	27%	5.4	16%	79%
Class Level	47.4	48%	8.8	25%	82%
Pupil Level	24.7	25%	20.7	59%	16%
Total (all levels combined)	98.0	100%	34.9	100%	64%

Table 4a: **Primary outcome = GL PiM13 Y8 - Subscale 2 (WEAK or STRONG MR ITEMS)**

Y8, All regions combined.

Intention to Treat Analyses

n=2,076 in 110 Y8 classes in 56 secondary schools

	Empty (Null) Model		**Control / Intervention Alone**		**Pupil & School Level Variables**		**Final Model**	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	-1.6	0.90	-0.2	0.41	0.0	0.44
Pre-test (KS2 Maths)	-	-	-	-	**3.1**	**0.14**	**3.2**	**0.13**
FSM (eligible & claiming)	-	-	-	-	-0.1	0.15	/	/
Gender (female)	-	-	-	-	**-0.3**	**0.13**	/	/
SEN (has SEN)	-	-	-	-	0.2	0.18	/	/
Age (in months)	-	-	-	-	0.0	0.02	/	/
School Level Attainment	-	-	-	-	-1.2	0.65	/	/
School level (selective	-	-	-	-	**3.3**	**0.94**	**3.8**	**0.97**
School level (%FSM)	-	-	-	-	0.0	0.03	/	/
School level (IDACI)	-	-	-	-	1.0	2.16	/	/
School level (OFSTED)	-	-	-	-	0.9	0.59	/	/
Constant	**6.7**	**0.47**	**7.5**	**0.64**	**12.5**	**2.77**	**6.9**	**0.32**

Effect sizes

Hedges g	-	-0.32	-0.04	+0.01
95% CIs		(-0.69, +0.05)	(-0.20, +0.13)	(-0.17, +0.18)

Table 4b: *Variance structure & explanatory power* - **GL PiM13 Y8 Subscale 2**

Intention to Treat Analysis

Y8 subsample - All regions combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	-5,285.8		-5,081.3		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level	5.8	24%	1.2	12%	79%
Class Level	10.8	44%	1.8	18%	83%
Pupil Level	7.9	32%	7.0	70%	11%
Total (all levels combined)	24.5	100%	10.1	100%	59%

Table 5a: **Primary outcome = GL PiM13 Y8 - Subscale 3 (STRONG MR ITEMS)**

Y8, All regions combined.

Intention to Treat Analyses

n=2,076 in 110 Y8 classes in 56 secondary schools

	Empty (Null) Model		**Control / Intervention Alone**		**Pupil & School Level Variables**		**Final Model**	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	-0.8	0.48	0.0	0.21	0.1	0.22
Pre-test (KS2 Maths)	-	-	-	-	**1.9**	**0.08**	**1.9**	**0.08**
FSM (eligible & claiming)	-	-	-	-	0.0	0.10	/	/
Gender (female)	-	-	-	-	**-0.2**	**0.08**	**-0.2**	**0.08**
SEN (has SEN)	-	-	-	-	0.0	0.12	/	/
Age (in months)	-	-	-	-	0.0	0.01	**-0.0**	**0.01**
School Level Attainment	-	-	-	-	**-0.8**	**0.33**	/	/
School level (selective	-	-	-	-	**1.5**	**0.48**	**1.6**	**0.51**
School level (%FSM)	-	-	-	-	0.0	0.01	**-0.0**	**0.01**
School level (IDACI)	-	-	-	-	0.1	1.10	/	/
School level (OFSTED)	-	-	-	-	0.6	0.30	/	/
Constant	**2.9**	**0.25**	**3.3**	**0.34**	**8.0**	**1.74**	**7.7**	**1.72**
Effect sizes								
Hedges g	-		-0.28		-0.01		+0.02	
95% CIs			(-0.62, +0.06)		(-0.16, +0.14)		(-0.14, +0.18)	

Table 5b: *Variance structure & explanatory power* **- GL PiM13 Y8 Subscale 3 Intention to Treat Analysis**

Y8 subsample - All regions combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	-4300.1		-4105.3		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level	1.7	21%	0.3	9%	81%
Class Level	2.9	38%	0.4	12%	86%
Pupil Level	3.1	41%	2.8	79%	10%
Total (all levels combined)	7.7	100%	3.5	100%	54%

Table 6a: **Primary outcome = GL PiM14 Y9 - Subscale 1 (ANY MR ITEMS)**

Y9, All regions combined.

Intention to Treat Analyses

n=2,101 in 111 Y9 classes in 52 secondary schools

	Empty (Null) Model		**Control / Intervention Alone**		**Pupil & School Level Variables**		**Final Model**	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	**-3.1**	**1.42**	-1.1	0.77	-1.3	0.78
Pre-test (KS2 Maths)	-	-	-	-	**5.0**	**0.25**	**4.9**	**0.25**
FSM (eligible & claiming)	-	-	-	-	-0.3	0.25	/	/
Gender (female)	-	-	-	-	-0.1	0.20	/	/
SEN (has SEN)	-	-	-	-	**-0.9**	**0.30**	**-0.8**	**0.30**
Age (in months)	-	-	-	-	0.0	0.03	/	/
School Level Attainment	-	-	-	-	-2.1	1.53	/	/
School level (selective	-	-	-	-	**10.8**	**1.86**	**11.5**	**0.58**
School level (%FSM)	-	-	-	-	-0.1	0.05	/	/
School level (IDACI)	-	-	-	-	0.9	3.70	/	/
School level (OFSTED)	-	-	-	-	2.2	1.43	/	/
Constant	**13.3**	**0.72**	**14.9**	**1.03**	**13.1**	**4.73**	**14.1**	**0.58**
Effect sizes								
Hedges g	-		**-0.37***		-0.13		-0.15	
95% CIs			**(-0.70, -0.04)**		(-0.31, +0.05)		(-0.33 +0.03)	

Table 6b: *Variance structure & explanatory power* - **GL PiM14 Y9 Subscale 1**

Intention to Treat Analysis

Y9 subsample - All regions combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	6,322.6		-6,143.6		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level	0	-	0	-	-
Class Level	56.7	74%	14.5	45%	74%
Pupil Level	19.6	26%	17.6	55%	10%
Total (all levels combined)	76.3	100%	32.1	100%	58%

Table 7a: **Primary outcome = GL PiM14 Y9 Subscale 2 (WEAK or STRONG MR ITEMS)**
Y9, All regions combined.
Intention to Treat Analyses

n=2,101 in 111 Y9 classes in 52 secondary schools

	Empty (Null) Model		Control / Intervention Alone		Pupil & School Level Variables		Final Model	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	**-0.8**	**0.38**	-0.2	0.17	-0.2	0.18
Pre-test (KS2 Maths)	-	-	-	-	**1.6**	**0.08**	**1.6**	**0.08**
FSM (eligible & claiming)	-	-	-	-	-0.1	0.09	/	/
Gender (female)	-	-	-	-	**-0.3**	**0.07**	**-0.3**	**0.07**
SEN (has SEN)	-	-	-	-	-0.1	0.11	/	/
Age (in months)	-	-	-	-	0.0	0.01	/	/
School Level Attainment	-	-	-	-	-0.6	0.34	/	/
School level (selective	-	-	-	-	**2.9**	**0.41**	**3.1**	**0.42**
School level (%FSM)	-	-	-	-	0.0	0.01	/	/
School level (IDACI)	-	-	-	-	0.4	0.82	/	/
School level (OFSTED)	-	-	-	-	0.5	0.32	/	/
Constant	**2.8**	**0.19**	**3.2**	**0.28**	**1.1**	**1.68**	**3.1**	**0.14**

Effect sizes

Hedges g	-	**-0.30***	-0.08	-0.09
95% CIs		**(-0.61, -0.001)**	(-0.21, +0.06)	(-0.23 +0.05)

Table 6b: ***Variance structure & explanatory power*** **- GL PiM14 Y9 Subscale 2**
Intention to Treat Analysis

Y9 subsample - All regions combined

	Empty (Null) Model		Final Model		
- 2 Log-likelihood	-4,118.3		-3,957.7		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level	0	-	0	-	-
Class Level	4.0	61%	0.7	22%	83%
Pupil Level	2.5	39%	2.3	78%	7%
Total (all levels combined)	6.4	100%	3.0	100%	54%

Table 8a: **Primary outcome = GL PiM14 Y9 Subscale 3 (STRONG MR ITEMS)**

Y9, All regions combined.

Intention to Treat Analyses

n=2,101 in 111 Y9 classes in 52 secondary schools

	Empty (Null) Model		**Control / Intervention Alone**		**Pupil & School Level Variables**		**Final Model**	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	-0.3	0.18	-0.1	0.08	-0.1	0.08
Pre-test (KS2 Maths)	-	-	-	-	**0.9**	**0.05**	**0.9**	**0.05**
FSM (eligible & claiming)	-	-	-	-	0.0	0.06	/	/
Gender (female)	-	-	-	-	**-0.2**	**0.05**	**-0.2**	**0.05**
SEN (has SEN)	-	-	-	-	-0.1	0.07	/	/
Age (in months)	-	-	-	-	0.0	0.01	/	/
School Level Attainment	-	-	-	-	-0.2	0.15	/	/
School level (selective	-	-	-	-	**1.1**	**0.18**	**1.2**	**0.18**
School level (%FSM)	-	-	-	-	0.0	0.00	/	/
School level (IDACI)	-	-	-	-	0.2	0.36	/	/
School level (OFSTED)	-	-	-	-	0.2	0.14	/	/
Constant	**1.3**	**0.09**	**1.5**	**0.13**	**0.7**	**1.07**	**1.5**	**0.06**
Effect sizes								
Hedges g	-		-0.25		-0.07		-0.08	
95% CIs			(-0.52, +0.01)		(-0.18, +0.04)		(-0.19 +0.04)	

Table 8b: *Variance structure & explanatory power* **- GL PiM14 Y9 Subscale 3**

Intention to Treat Analysis

Y9 subsample - All regions combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	-3122.6		-2989.2		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level	0	-	0	-	-
Class Level	0.8	46%	0.1	9%	89%
Pupil Level	1.0	55%	1.0	91%	4%
Total (all levels combined)	1.8	100%	1.1	100%	42%

D4: Maths Attainment (GL PiM 12 to 14 Standardised Age Score) models (On Treatment Analyses).

(Summarised in Section 7.6 in the main report)

NOTE - to assist in reading these tables, please refer to Appendix E1 (glossary & effect sizes)

Table 1a: Primary outcome = GL PiM12to13 Combined Measure (all areas combined)

On-treatment Analyses

n=5,253 pupils in 269 classes in 56 secondary schools

	Empty (Null) Model		**Control / Intervention Alone**		**Pupil & School Level Variables**		**Final Model**	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	**-2.5**	**2.35**	-0.2	0.73	-0.1	0.74
Pre-test (KS2 Maths)	-	-	-	-	**11.4**	**0.23**	**11.4**	**0.24**
FSM (eligible & claiming)	-	-	-	-	**-1.3**	**0.28**	**-1.3**	**0.28**
Gender (female)	-	-	-	-	**0.7**	**0.22**	**0.7**	**0.22**
SEN (has SEN)	-	-	-	-	**-1.9**	**0.34**	**-1.9**	**0.34**
School Level Attainment	-	-	-	-	**-2.4**	**1.22**	**-2.4**	**1.23**
School level (selective	-	-	-	-	**10.5**	**1.59**	**10.6**	**1.60**
School level (%FSM)	-	-	-	-	**-0.1**	**0.05**	**-0.1**	**0.03**
School level (IDACI)	-	-	-	-	1.6	3.64	/	/
School level (OFSTED)	-	-	-	-	**2.9**	**1.09**	**2.9**	**1.10**
Constant	**98.3**	**1.19**	**99.5**	**1.66**	**99.6**	**1.27**	**99.6**	**1.27**

Effect sizes

Hedges g	-	-0.17	-0.01	-0.01
95% CIs		(-0.47, +0.14)	(-0.11, +0.09)	(-0.10, +0.09)

Table 1b: ***Variance structure & explanatory power* - GL PiM12to13 Combined Measure**

Intention to Treat Analysis (Overall Maths Attainment)

All Regions / Year Groups combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	-19082.9		-18208.5		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level Variance	50.5	21%	2.8	4%	95%
Class Level Variance	119.5	50%	15.8	22%	87%
Pupil Level Variance	69.4	29%	54.4	75%	22%
Total (all levels combined)	239.4	100%	72.9	100%	70%

Table 2a: **Primary outcome = GL PiM12 Y7 (all areas combined)**

On Treatment Analyses

n=1,844 pupils in 97 Y7 classes in 49 secondary schools

	Empty (Null) Model		**Control / Intervention Alone**		**All Pupil & School Level Variables**		**Final Model**	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	-2.0	3.04	-0.1	0.85	0.0	0.89
Pre-test (KS2 Maths)	-	-	-	-	**13.0**	**0.33**	**13.0**	**0.33**
FSM (eligible & claiming)	-	-	-	-	**-1.9**	**0.47**	**-1.9**	**0.47**
Gender (female)	-	-	-	-	**1.2**	**0.38**	**1.2**	**0.38**
SEN (has SEN)	-	-	-	-	**-2.5**	**0.58**	**-2.5**	**0.58**
School Level Attainment	-	-	-	-	**-3.1**	**1.39**	/	/
School level (selective	-	-	-	-	**6.3**	**1.77**	**6.3**	**1.86**
School level (%FSM)	-	-	-	-	-0.1	0.05	**-0.1**	**0.04**
School level (IDACI)	-	-	-	-	0.3	4.03	/	/
School level (OFSTED)	-	-	-	-	1.6	1.23	/	/
Constant	**99.8**	**1.53**	**100.8**	**2.09**	**102.6**	**1.46**	**101.1**	**0.88**
Effect sizes								
Hedges g	-		-0.13		0.00		0.00	
95% CIs			(-0.52, +0.26)		(-0.11, +0.11)		(-0.11, +0.11)	

Table 2b: *Variance structure & explanatory power -* **GL PiM12 Y7**

On Treatment Analysis (Overall Maths Attainment)

Y7 subsample - All regions combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	-6893.4		-6376.4		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level	44.2	18%	4.2	7%	91%
Class Level	121.5	48%	4.9	8%	96%
Pupil Level	86.1	34%	55.1	86%	36%
Total (all levels combined)	251.7	100%	64.2	100%	75%

Table 3a: Primary outcome = GL PiM13 Y8 (all areas combined)

On Treatment Analyses

n=1,761 pupils in 89 Y8 classes in 52 secondary schools

	Empty (Null) Model		**Control / Intervention Alone**		**Pupil & School Level Variables**		**Final Model**	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	-3.0	3.43	0.0	1.35	0.2	1.36
Pre-test (KS2 Maths)	-	-	-	-	**10.2**	**0.44**	**10.1**	**0.43**
FSM (eligible & claiming)	-	-	-	-	**-1.0**	**0.48**	**-1.0**	**0.48**
Gender (female)	-	-	-	-	0.3	0.38	/	/
SEN (has SEN)	-	-	-	-	**-2.1**	**0.57**	**-2.2**	**0.56**
School Level Attainment	-	-	-	-	-1.4	2.17	/	/
School level (selective	-	-	-	-	**11.7**	**2.88**	**12.6**	**2.87**
School level (%FSM)	-	-	-	-	-0.1	0.08	/	/
School level (IDACI)	-	-	-	-	1.9	7.06	/	/
School level (OFSTED)	-	-	-	-	**3.8**	**1.94**	**3.8**	**1.58**
Constant	**98.5**	**1.73**	**100.0**	**2.36**	**98.3**	**2.33**	**96.3**	**1.52**
Effect sizes								
Hedges g	-		-0.20		0.00		+0.02	
95% CIs			(-0.63, +0.24)		(-0.17, +0.17)		(-0.16, +0.19)	

Table 3b: Variance structure & explanatory power - **GL PiM13 Y8**

On Treatment Analysis (Overall Maths Attainment)

Y8 subsample - All regions combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	-6345.8		-6120.2		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level	83.1	33%	10.1	12%	88%
Class Level	107.0	42%	16.2	20%	85%
Pupil Level	64.9	26%	54.7	67%	16%
Total (all levels combined)	255.0	100%	80.9	100%	68%

Table 4a: Primary outcome = GL PiM14 Y9 (all areas combined)

On Treatment Analyses

n=1,648 pupils in 83 Y9 classes in 45 secondary schools

	Empty (Null) Model		**Control / Intervention Alone**		**All Pupil & School Level Variables**		**Final Model**	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Group (1= Intervention)	-	-	-3.2	2.95	0.0	1.40	0.0	1.50
Pre-test (KS2 Maths)	-	-	-	-	**8.8**	**0.48**	**8.8**	**0.48**
FSM (eligible & claiming)	-	-	-	-	-0.8	0.48	/	/
Gender (female)	-	-	-	-	0.3	0.38	/	/
SEN (has SEN)	-	-	-	-	-1.0	0.57	/	/
School Level Attainment	-	-	-	-	-4.0	2.74	/	/
School level (selective	-	-	-	-	**16.8**	**3.00**	**18.1**	**3.13**
School level (%FSM)	-	-	-	-	-0.1	0.08	/	/
School level (IDACI)	-	-	-	-	7.4	6.79	/	/
School level (OFSTED)	-	-	-	-	4.4	2.47	/	/
Constant	**96.6**	**1.49**	**98.2**	**2.03**	**96.9**	**2.42**	**96.1**	**1.06**

Effect sizes

Hedges g	-	-0.23	0.00	0.00
95% CIs		(-0.64, +0.18)	(-0.19, +0.20)	(-0.21, +0.21)

Table 4b: ***Variance structure & explanatory power* - GL PiM14 Y9**

On Treatment Analysis (Overall Maths Attainment)

Y9 subsample - All regions combined

	Empty (Null) Model		**Final Model**		
- 2 Log-likelihood	-5814.9		-5677.3		
	Variance Remaining	ICC (as %)	Variance Remaining	ICC (as %)	**R^2 (Explanatory Power)**
School Level	13.0	6%	1.4	2%	89%
Class Level	153.0	69%	37.2	42%	76%
Pupil Level	55.6	25%	50.2	57%	10%
Total (all levels combined)	221.6	100%	88.9	100%	60%

D5: MRP Randomisation Notes

To begin with, the final numbers of schools per area and some detail on the schools that were excluded prior to randomisation are summarised:

Region 1	21 schools in final sample
Region 2	19 in final sample
Region 3	22 in final sample
All Regions Combined	62 schools in final sample - 2 schools dropped out immediately following randomisation.

In each area, 10 schools will be randomly selected for the MRP intervention. This was done using a stratification scheme based on geography and school level (GCSE) attainment.

Within each of the three areas, randomisation was done by stratifying on attainment - and using the national average of 59% attaining 5+ A*-C (taken from Gov.uk website).

Schools where a greater proportion attained this level (i.e. 60% or higher) were placed in the 'higher attaining' group whilst schools where the proportion attaining this level was 59% or lower were placed in the lower attaining group[42].

The result, in terms of numbers is illustrated in the table below. The first number represents the total number of schools in each strata whilst the second one in brackets and red indicates the number of schools that will be randomly assigned to become intervention schools.

Area	Lower Attaining	Higher Attaining
Region 1	6 **(3)**	15 **(7)**
Region 2	4 **(2)**	15 **(8)**
Region 3	4 **(2)**	18 **(8)**

Randomisation was done using SPSS - the SYNTAX and guidance notes are included as an appendix for this document.

APPENDIX: SPSS SYNTAX used for MRP Randomisation

[42] This was done for all schools with GCSE attainment details. There was an instance when this data was not available and so judgement was used. In Region 1, a Middle School was placed in the higher attaining group based on FSM and IDACI profile.

*   For Region 1 LOWER ATTAINER Strata ...

FILTER OFF.

USE ALL.

SELECT IF (MRP_Strata=1.1).

EXECUTE.

USE ALL.

SAMPLE 3 from 6.

EXECUTE.

*   For Region 1 HIGHER ATTAINER Strata ...

FILTER OFF.

USE ALL.

SELECT IF (MRP_Strata=1.2).

EXECUTE.

USE ALL.

SAMPLE 7 from 15.

EXECUTE.

*   For Region 2 LOWER ATTAINER Strata ...

FILTER OFF.

USE ALL.

SELECT IF (MRP_Strata=2.1).

EXECUTE.

USE ALL.

SAMPLE 2 from 4.

EXECUTE.

*   For Region 2 HIGHER ATTAINER Strata ...

FILTER OFF.

USE ALL.

SELECT IF (MRP_Strata=2.2).

EXECUTE.

USE ALL.

SAMPLE 8 from 15.

EXECUTE.

*   For Region 3 LOWER ATTAINER Strata ...

FILTER OFF.

USE ALL.

SELECT IF (MRP_Strata=3.1).

EXECUTE.

USE ALL.

SAMPLE 2 from 4.

EXECUTE.

*   For Region 3 HIGHER ATTAINER Strata ...

FILTER OFF.

USE ALL.

SELECT IF (MRP_Strata=3.2).

EXECUTE.

USE ALL.

SAMPLE 8 from 18.

EXECUTE.

Department for Education

Reference: DFE- RR406B

ISBN: 978-1-78105-487-1

This research was commissioned under the 2010 to 2015 Conservative and Liberal Democrat coalition government. Views expressed in this report are those of the authors. They do not necessarily reflect government policy.

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

Any enquiries regarding this publication should be sent to us at: Christopher.PRICE@education.gsi.gov.uk or www.education.gov.uk/contactus

This document is available for download at www.gov.uk/government/publications