Trial Evaluation Protocol
Maths-for-Life

Evaluator: Behavioural Insights Team
Principal investigator: Michael Sanders

PROJECT TITLE	Maths-for-Life
DEVELOPER	University of Nottingham
EVALUATOR	Behavioural Insights Team
PRINCIPAL INVESTIGATOR	Michael Sanders
PROTOCOL AUTHOR(S)	David Nolan, Patrick Taylor, Pantelis Solomon, Jessica Heal
TRIAL DESIGN	Two-arm cluster randomised controlled trial with random allocation at setting (FE college/6th Form College/School/Training Provider) level
AGE RANGE	16+, KS5 GCSE Maths resit students
NUMBER OF SCHOOLS	Approx 110 Settings (50 FE Colleges & 60 Schools)
NUMBER OF PUPILS	Approx 4600 (Assuming 80 students per college, 10 students per school)
PRIMARY OUTCOME	GCSE Maths Attainment Uniform Mark Scheme (UMS) scores

Protocol version history

VERSION	DATE	REASON FOR REVISION
1.1		
1.0	19 July 2018	N/A

Table of Contents

Intervention

This intervention aims to improve GCSE mathematics retake outcomes for post-16 students (KS5) in Further Education Colleges, 6th Form Colleges, schools and training providers. It attempts to develop a more student-centred classroom approach based on problem solving and dialogic teaching. Dialogic teaching aims to simulate learning through classroom conversation, both peer-to-peer and teacher-to-student. The intervention will be delivered by the Centre for Research in Mathematics Education at the University of Nottingham. It builds on an evidence-based corpus of classroom materials, using resources from the Standards Unit Box (Improving Learning in Mathematics) and other resources developed over by Malcolm Swan and colleagues[1,2].

The focus is on five key areas of the mathematics curriculum that are known to be challenging and present difficulties to GCSE mathematics students, these are: proportional reasoning; algebraic expressions; parts of a whole; contextual problems; and handling data. The materials address these key mathematical areas and concepts using contexts and problems designed to re-engage GCSE resit students in mathematics. Many of these students experience disaffection and disengagement after "failure" (achieving a grade 3 or below) in their KS4 GCSE exam. The intervention addresses this by introducing a problem-solving approach, adopting a student-centred focus, using discussion, and using research-informed diagnostic and formative assessment. Tasks are designed to be used with students working collaboratively. For example, students could be given a set of cards with different objects on (a tall skyscraper, the length of a fly, the distance to the moon) and asked to work in groups to match each object to their corresponding measurement.

The intervention supports teachers by providing evidence-informed materials together with a professional development (PD) programme based on Wake and Swan's lesson study research[3]. The PD takes an "action research" approach, led by a cadre of trained teacher PD Leads in which teacher research groups engage in five cycles of classroom-based inquiry into effective pedagogies, supported by an online toolkit. The programme aims to address a skill shortage amongst teachers and attempts to change how mathematics is conceptualised by young people, moving from a binary subject where thinking is 'right' or 'wrong' to one that is debated and discussed.

The intervention begins and ends with an event for Class Teachers. In between these two meetings, the intervention goes through the following lesson study cycle:

1. Clusters of Class Teachers meet to learn about and plan a Maths-for-Life lesson, supported by their Lead Teacher.
2. Class Teachers deliver a Maths-for-Life lesson to their own class.
3. Class Teachers meet as a cluster to observe a peer from the cluster delivering the same lesson.
4. Clusters meet again with their Lead Teacher to reflect on the lesson delivered, and to learn about the next lesson to be delivered.

---

[1] Swan, M. (2006). Learning GCSE mathematics through discussion: what are the effects on students?. Journal of Further and Higher Education, 30(3), 229-241.

[2] Swan, M. (2007). The impact of task-based professional development on teachers' practices and beliefs: A design research study. Journal of Mathematics Teacher Education, 10(4-6), 217-237.

[3] Swan, M., & Swain, J. (2010). The impact of a professional development programme on the practices and beliefs of numeracy teachers. Journal of further and Higher Education, 34(2), 165-177.

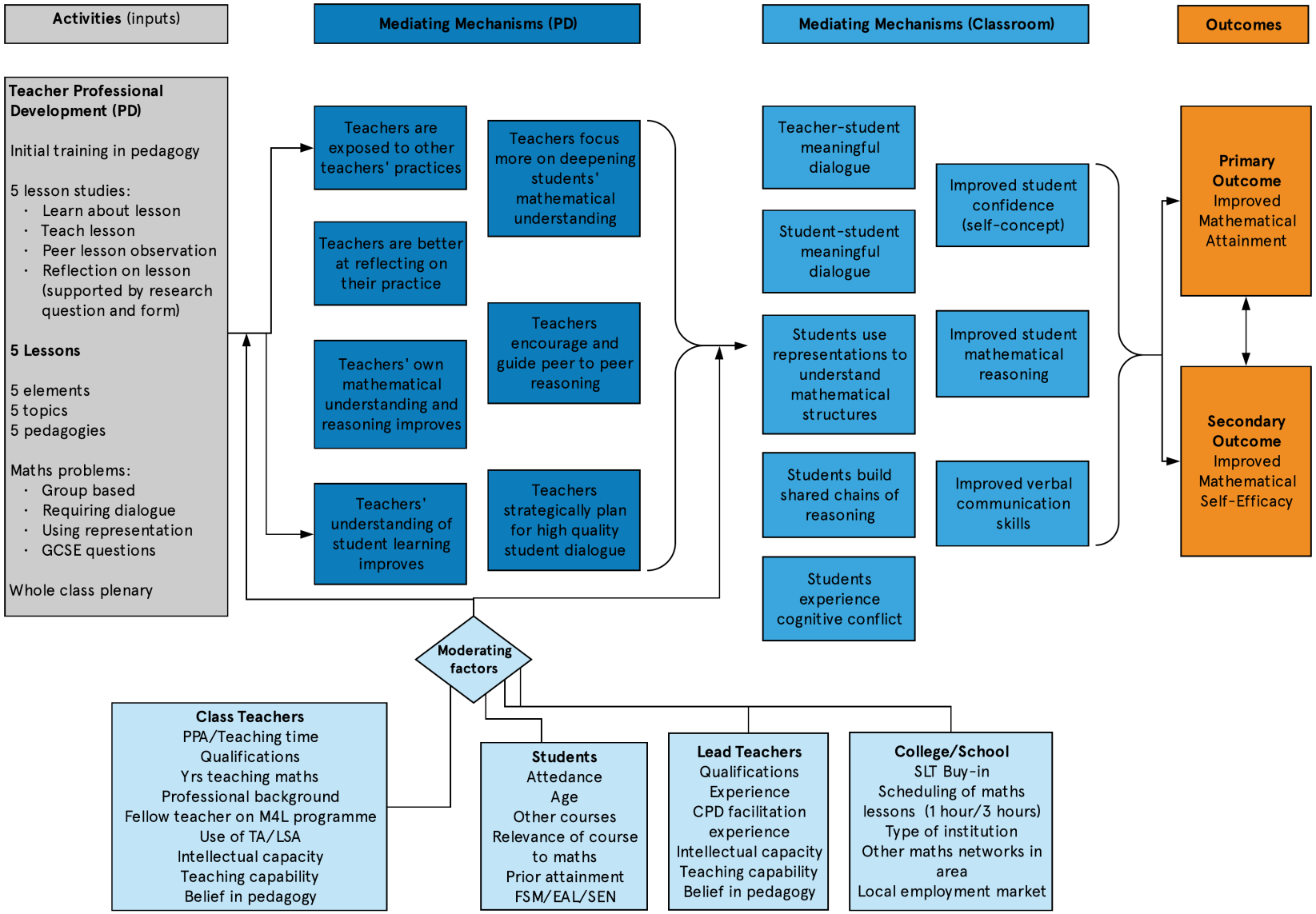
This cycle is completed five times, with a new lesson being delivered and studied each time. The five lessons are delivered between November and April. Each lesson is designed to last for one hour. The PD requires a total of 6 days of teacher time and breaks down as follows:

- Launch event: 0.5 days
- Lesson planning and reflection: 1 day per lesson (5 days in total)
- Closing event: 0.5 days

Lessons are taught in regular classrooms in schools and colleges. Cluster PD sessions take place in a range of regional locations, such as participating schools and colleges. For a more detailed outline of the intervention please refer to the TiDieR framework in Appendix 1. The programme has been piloted with 20 lead teachers, across 20 colleges, in the academic year 2017-18. In the 2018-19 academic year, these 20 teachers will support a cohort of approximately 5 new teachers each through the PD programme. This process will see approximately 100 new teachers applying the Maths-for-Life pedagogy with their GCSE mathematics resit classes.

Logic Model

An IDEA workshop was held, using the TIDieR framework, to develop a logic model (see over) in collaboration with the developers at the University of Nottingham. The Logic Model will be instrumental in directing the impact evaluation and implementation and process evaluation (IPE). Throughout the IPE, we will attempt to monitor the proposed mediating mechanisms as well as understand the role played by potential moderators. The TIDieR framework for the intervention can be found in Appendix 1.


Activities (inputs)
Mediating Mechanisms (PD)
Mediating Mechanisms (Classroom)
Outcomes
Teacher Professional Development (PD)
Initial training in pedagogy
5 lesson studies:
· Learn about lesson
· Teach lesson
· Peer lesson observation
· Reflection on lesson (supported by research question and form)
5 Lessons
5 elements
5 topics
5 pedagogies
Maths problems:
· Group based
· Requiring dialogue
· Using representation
· GCSE questions
Whole class plenary
Teachers are exposed to other teachers' practices
Teachers are better at reflecting on their practice
Teachers' own mathematical understanding and reasoning improves
Teachers' understanding of student learning improves
Teachers focus more on deepening students' mathematical understanding
Teachers encourage and guide peer to peer reasoning
Teachers strategically plan for high quality student dialogue
Teacher-student meaningful dialogue
Student-student meaningful dialogue
Students use representations to understand mathematical structures
Students build shared chains of reasoning
Students experience cognitive conflict
Improved student confidence (self-concept)
Improved student mathematical reasoning
Improved verbal communication skills
Primary Outcome
Improved Mathematical Attainment
Secondary Outcome
Improved Mathematical Self-Efficacy
Moderating factors
Class Teachers
PPA/Teaching time
Qualifications
Yrs teaching maths
Professional background
Fellow teacher on M4L programme
Use of TA/LSA
Intellectual capacity
Teaching capability
Belief in pedagogy
Students
Attedance
Age
Other courses
Relevance of course to maths
Prior attainment
FSM/EAL/SEN
Lead Teachers
Qualifications
Experience
CPD facilitation experience
Intellectual capacity
Teaching capability
Belief in pedagogy
College/School
SLT Buy-in
Scheduling of maths lessons (1 hour/3 hours)
Type of institution
Other maths networks in area
Local employment market

Significance

Around 30% of young people failed to attain a grade 4 in maths GCSE during year 11 in 2017[4]. A significant proportion of these young people now retake this exam in Further Education or 6th Form colleges, with the remainder retaking in school sixth-forms. However, in prior years only 29.5% and 28% of those resitting their mathematics GCSE have obtained a pass[5][6]. Those students who do not obtain a pass grade must then continue resitting until they pass or turn 19. Mathematical literacy is a critical life skill. Despite its economic and social importance, numeracy in the UK is weak. According to the OECD, one-third of 16-19-year olds in the UK have poor mathematical skills, three times as many as the highest performing countries[7]. For the individual, achieving a pass in GCSE mathematics opens doors to better longer term educational and employment outcomes[8]

The evidence for the principles behind the programme is best described in a number of publications by Malcolm Swan, the lead designer and researcher of the materials on which the intervention is based[9]. Fundamentally, the teaching resources draw on design principles from diagnostic teaching research. Swan reports evidence that effective use of the materials in student centred ways in post-16 contexts leads to increases in attainment. One study used a pre/post-test design to assess the outcomes of students (N=334) who received 'many' or 'few' of the lessons. Those who received 'many' had statistically significant gains on an algebra test compared to their peers who received 'few'. In another study, materials were adapted for the US and were evaluated using a quasi-experimental design (N=471) with a matched control group. This later study found that the intervention group made significant gains in attainment compared to the control group (0.13 Cohen's d effect size, which is equivalent to 4.6 months of schooling)[10].

A pilot evaluation was carried out in advance of this trial. In the 2017-18 academic year, the University of Nottingham ran a pilot version of the Maths-for-Life programme with 20 teachers from 20 colleges, with the Behavioural Insights Team (BIT) acting as the independent evaluator. This study aimed to evaluate the promise, feasibility and readiness for trial of the programme, using a combination of qualitative cases studies of pilot colleges, observations of training days, quantitative surveying, and interviews with 4 non-pilot colleges. The findings suggested that this is an ambitious project, due to the very low levels of student confidence, the difficult-to-master pedagogy, and the likely dilution of quality in delivery as the programme scales. However, we also found evidence of promise in terms of student outcomes, and a confidence amongst pilot colleges that the programme is feasible to deliver and ready for trial. The details of these findings have influenced the refinement of the intervention (for example, considering the role of Teaching Assistants), its logic model and the focus of the implementation and process evaluation (for example, paying particular

---

[4] Murray, C. (2017). English and Maths GCSE resit results 2017. *FE Week*. Available online: https://feweek.co.uk/2017/08/24/english-and-maths-gcse-resit-results-2017/. [Last accessed: 30 April 2018].

[5] Allen, B. (2016). Repeat after 'E': the treadmill of post-16 GCSE maths and English retakes. Education Datalab. Available online: https://educationdatalab.org.uk/2016/08/repeat-after-e-the-treadmill-of-post-16-gcse-maths-and-english-retakes/. [Last accessed: 30 April 2018]

[6] Thomson, D. (2017) GCSE results day 2017: Good news about resits in English. Education Datalab. Available online: https://educationdatalab.org.uk/2017/08/gcse-results-day-2017-good-news-about-resits/ [Last accessed: 30 April 2018]

[7] Kuczera, M., Field, S., & Windisch, H. C. (2016). Building skills for all: a review of England. Policy Insights from the Survey of Adult Skills. Paris, France: OECD.

[8] Hayward, H., Hunt, E., Lord, A., Vernoit, J., North, C., & Donnelly, E. (2014). The economic value of key intermediate qualifications: estimating the returns and lifetime productivity gains to GCSEs, A levels and apprenticeships.

[9] Malcolm Swan (2007) Learning GCSE mathematics through discussion: what are the effects on students?, Journal of Further and Higher Education, 30:3, 229-241, DOI: 10.1080/03098770600802263.

[10] Herman, J. L., Matrundola, D. L. T., Epstein, S., Leon, S., Dai, Y., Reber, S., & Choi, K. (2015). The Implementation and Effects of the Mathematics Design Collaborative (MDC): Early Findings from Kentucky Ninth-Grade Algebra 1 Courses. CRESST Report 845. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*.

attention to the variation in competency of the Class Teachers and Lead Teachers as the programme is scaled up).

Impact evaluation methodology

Research questions

The primary research question of this evaluation is the following: Does the Maths-for-Life programme improves pupils' overall GCSE mathematics resit performance in KS5, as measured by the moderated UMS point scores?

The evaluation will also address the following questions as secondary objectives:

- What is the impact of the programme on KS5 GCSE mathematics resit pass-rates, as measured by the percentage of students achieving a grade 4 or above?
- What is the impact of the programme on students' mathematical self-efficacy, as measured by Part E of the Year 10 Teleprism survey[11]?

The intervention will be delivered in the academic year 2018/19. BIT will assess the impact of the programme on the student outcomes of KS5 GCSE mathematics performance as measured by the moderated UMS point scores, and KS5 GCSE mathematics resit pass-rates, for two cohorts of students:

- **Cohort 1**: Students resitting GCSE mathematics in Summer 2019 (Academic year 18/19)
- **Cohort 2**: Students resitting GCSE mathematics in Summer 2020 (Academic year 19/20)

This is to ensure that the evaluation captures any longer-term impact of the intervention, given changes on teacher practice may take time to embed.

Design

This will be a two-arm randomised controlled trial (efficacy), with randomisation at the setting (FE college/6th Form College/school) level, stratified by setting type (FE college or 6th Form College/School/Training Provider). The two arms are:

1. The treatment arm, in which nominated teachers from settings assigned to this arm receive the Maths-for-Life PD programme.
2. The control arm, in which settings continue as they otherwise would have.

The trial design is set out in the table below.

Trial type and number of arms	Two-arm randomised controlled trial
Unit of randomisation	Setting (FE/6th Form College/School/Training Provider) level
Stratification variables (if applicable)	Setting type – 3 strata (FE College, 6th Form College/School and Training Providers)[12]

[11] For information on the design and validation of this surveying approach see: Pampaka, M., Kleanthous, I., Hutcheson, G. D., & Wake, G. (2011). Measuring mathematics self-efficacy as a learning outcome. Research in Mathematics Education, 13(2), 169-190. The survey itself is available online at: http://www.teleprism.com/surveys.htm. [Last accessed: 30 April 2018].

[12] Stratification will group 6th Form Colleges, Schools and Training Providers to form one stratum to ensure equal numbers of students in treatment and control given they are likely to be aligned in numbers of classes teachers and students

Outcomes	primary	KS5 GCSE mathematics UMS score
	secondary	KS5 GCSE mathematics pass rate Student self-reported mathematical self-efficacy
Outcome sources **(instruments, datasets)**	primary	Data will be collected directly from colleges
	secondary	National Pupil Database (NPD) Administered survey

The trial will be complemented by an implementation and process evaluation (IPE) that will focus on the following domains:

- **Mechanisms:** To contribute to our understanding of how the intervention works
- **Fidelity:** To assess the extent to which the intervention is delivered as intended
- **Context:** To assess how participant characteristics and the implementation context relate to the effectiveness of the treatment

Randomisation

Settings will be randomised into trial arms in October 2018 following the recruitment process. Randomisation will be conducted using the data analysis and statistical software Stata 14. The code used to carry out this randomisation will be recorded and reported in the final report.

Randomisation will be stratified on setting type to ensure the ratio of colleges to schools is equal in both arms. This will be done for two reasons:

1. To ensure the total number of students is close to equal across both trial arms, given the number of students varies substantially between colleges and schools. This will maximise our statistical power.
2. To prevent any bias in our analysis that may arise from systematic differences between FE colleges, 6th Form Colleges/Schools and Training Providers.

The intervention requires clusters of Class Teachers to meet regularly with their Lead Teacher. Therefore, with the aim of minimising travel for those participating, we present two strategies for accounting for the geographical location of participating settings and lead teachers.

1. Our preferred option is to randomise settings without stratifying on a geographical variable such as region. Afterwards, we will form groups of Class Teachers and Lead Teachers based on their geographical locations (from those allocated to the treatment group). That is, any given Class Teacher will be grouped together with the next four closest Class Teachers, alongside the most convenient Teacher. This is our preferred option as we believe it reflects how the process would occur in the absence of a trial. Additionally, stratifying on region may result in small strata, therefore weakening our randomisation strategy and ability to make causal claims.
2. If the above is not possible, we will stratify the randomisation on a regional variable to ensure allocation to groups is evenly distributed across regions. It must be noted that

this strategy may suffer from the issue of weak randomisation if the strata are very small.

To mitigate against the risk of control group setting attrition, randomisation and trial-arm allocation will only be communicated to settings after pupil data has been provided to BIT. All settings allocated to the control group will be financially reimbursed for their continued participation in the trial.

Participants

The trial will be conducted with KS5 mathematics re-sit students from recruited schools, 6th Form colleges, FE colleges, and Training Providers that are in the classes of teachers from each setting that are participating in the programme.

Settings will be recruited by the University of Nottingham. For settings to be eligible, they must have students that are retaking the mathematics GCSE in KS5. During recruitment, teachers will be selected by their settings to be part of the programme. The University of Nottingham expects to recruit one teacher per school, and one to two teachers per college. The eligibility criteria for teachers participating is as follows:

- **FE Colleges:** Teachers should be expecting to teach at least 80 students (if multiple teachers are recruited then this figure is a combined total).
- **Schools, 6th Form Colleges and Training Providers:** Teachers should be expecting to teach at least 10 students.

In the case of large FE colleges (or college groups) the teachers may be geographically and organisationally separate. For the purposes of the trial, these teachers will be considered to be teaching in distinct settings, despite coming from the same institutional group. As such, one large college could be classed as two settings for the purposes of recruitment, randomisation and analysis.

The programme will only accept more teachers from each setting than specified above if the programme has space to accommodate these teachers without limiting the capacity for more settings to participate.

Settings wishing to participate in the trial will be asked to sign a Memorandum of Understanding (MOU) (see Appendix 2) before enrolment in Spring 2018, agreeing to the required activities for both the intervention and evaluation. This includes the requirement for each school to collect and provide UMS scores for all participating students. Recruited settings must be satisfied that there is a clear *legitimate interests* argument to justify the sharing of the relevant student data with BIT (so consent from students and parents will not be necessary), as outlined in the ethical approval section below. Settings will also have to sign a Data Processing Agreement (DPA) with BIT.

Sample size calculations

To conduct all power analyses, we have used a statistical process known as simulation-based inference using the statistical software, R.

For simple trials (such as individually randomised controlled trials with no covariates and a binary outcome) we can use closed-form mathematical expressions to derive the power of a given experimental design. For more complex experiments, or where the assumptions underlying these formulae are not met, we simulate thousands of hypothetical experiments with a given effect size and observe how many of them return a significant result, which is known as a Monte Carlo measure of the experimental power. In this case, the combination

of setting randomisation and the presence of two 'types' of setting with quite different properties necessitates taking this approach.

We present the required sample sizes to achieve 80% statistical power for range of given Minimum Detectable Effect Sizes (MDESs) for our primary outcome measure, KS5 GGSE resit performance, as measured by UMS scores. We present our MDES in terms of Hedges' g. The MDES associated with our target number of recruited settings (50 FE colleges and 60 Schools) is presented in bold. The following assumptions were made:

- Statistical power: 80%
- Number of FE colleges: 50
- Students per FE college: 80[13]
- Number of schools: 60
- Students per school 6th form college/School/Training Provider: 10
- Intracluster Correlation Coefficient (ICC): 0.2 (Based on previous BIT FE trials[14])
- Student attrition: 20% (in line with EEF guidelines)
- Baseline measures (KS4 mathematics grade & KS2 mathematics score) have correlation 0.5 with our KS5 mathematics GCSE outcome.[15]
- Baseline resit pass-rate: 28%[16]
- KS4 GCSE mathematics outcomes follow a normal distribution with a mean score of 50, and standard-deviation of 20. This assumption is based on a number of sources, namely: Ofqual data on distribution of numerical grades (1-9)[17], the mathematics KS4 GCSE pass-rate of 73%[18] and the following average raw-score grade boundaries: Grade 4 - 33, Grade 3 - 17, Grade 2 - 10[19] (please see attached document in the footnote for full range of scores and grades) .
- The simulated treatment effect is uniform across all participants in the treatment group. Although this assumption is simplistic and may be unrealistic, there is not, ex ante, any empirical reason to assume any particular different functional form for the treatment effect.
- Percentage of FSM Students in Schools: 20% (in line with EEF Guidelines)
- Percentage of FSM Students in Colleges: 30%[20]

	OVERALL	FSM
MDES	0.15	0.22

[13] Estimate based on past BIT FE trials, additional anecdotal evidence provided by the University of Nottingham suggested this was a conservative estimate

[14] Hume, S et al. (2018). Retention and success in Maths and English: A Practitioner Guide to Applying Behavioural Insights

[15] Given the literature in this domain is quite sparse, there were no formal correlations available to use. We have made a conservative estimate revising down the KS2 and KS4 mathematics examination correlations from 0.7 (the correlation between KS2 & KS4 mathematics examinations) to 0.5 to account for this.

[16] Thomson, D. (2017) GCSE results day 2017: Good news about resits in English. Education Datalab. Available online: https://educationdatalab.org.uk/2017/08/gcse-results-day-2017-good-news-about-resits/ [Last accessed: 30 April 2018]

[17] Ofqual Analytics. (2017). Grade distributions of reformed (9-1) GCSEs. Available online: https://analytics.ofqual.gov.uk/apps/2017/GCSE/9to1/ [Last accessed: 30 April 2018]

[18] Murray, C. (2017). Exam Results: GCSE 2017 Results Maths. Schools week. Available online: https://schoolsweek.co.uk/gcse-results-2017-maths/ [Last accessed: 30 April 2018]

[19] Pearson Edexcel. (2015). Grade Boundaries Edexcel GCSE June 2015. Available online: https://qualifications.pearson.com/content/dam/pdf/Support/Grade-boundaries/GCSE/1506-GCSE-Grade-Boundaries.pdf [Last accessed: 30 April 2018]

[20] A conservative estimate revised upwards from the standard of 20% given FE College students are drawn from disproportionately disadvantaged backgrounds

Pre-test/ post-test correlations	level 1 (student)	0.5	0.5
Intracluster correlations (ICCs)	level 2 (setting)	0.2	0.2
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided
Average cluster size		80 students per college, 10 per school	24 students per college, 2 per school
Number of colleges	intervention	25	25
	control	25	25
	total	50	50
Number of schools	intervention	30	30
	control	30	30
	total	60	60
Number of pupils	intervention	2300	660
	control	2300	660
	total	4600	1320

Large variation in cluster size tends to reduce statistical power in cluster randomised controlled trials[21]. Often, researchers will try to factor this into their power analysis by an inflation factor known as the coefficient of variation, and incorporating this into the closed-form mathematical equations which are commonly used to derive statistical power for simple trials. However, given we are using a simulations-based approach, the effect of the variation in cluster size is accounted for as we are replicating the experiment thousands of times, and therefore any MDES estimates incorporate the issues presented by varying cluster sizes, as for each simulation we generate a hypothetical dataset which will have these properties.

Outcome Measures

The primary outcome is overall KS5 mathematics GCSE resit performance for the academic year 2018/2019, as per the UMS score collected and provided to BIT from the settings directly as specified in the MOU. This assessment will take place in summer 2019 and 2020.

The evaluation will also consider two secondary outcome measures:

- **KS5 mathematics GCSE pass rates,** as measured by a binary outcome variable indicating whether or not a student achieved a grade 4 or higher. This will be obtained from the NPD using the KS4_L2BASICS_94[22] variable (Achieved standard 9-4 passes in English and Maths GCSEs at Level 2), in December 2019. NPD data will be matched to original pupil data post-randomisation in December 2018 to obtain the baseline measures for each participating student.

21 Lauer, S. A., Kleinman, K. P., & Reich, N. G. (2015). The effect of cluster size variability on statistical power in cluster-randomized trials. PloS one, 10(4), e0119074.

22 The KS5 resit examination overwrites the KS4 Maths grade in the NPD tables

- **Student self-efficacy,** as measured by Part E of the Year 10 Teleprism survey[23]. Self-efficacy is the expectation an individual has that they will succeed in a task[24], is known to be malleable and can influence academic attainment.[25] Mathematical self-efficacy has been identified by the developers as the key intermediate outcome in this intervention that supports attainment. Findings from the pilot evaluation also suggested it to be a significant outcome. The Teleprism survey asks students to rate how confident they are - on a four-point Likert scale that ranges from 'Not confident at all' to 'Very confident' - in answering a range of questions that correspond to the following GCSE Maths topics: Number, Algebra, Geometry and measures, Ratio, proportion and rates of change, and Statistics. Although a four-point Likert scale may have low variance, and therefore low sensitivity in analysis, we believe this drawback is counteracted by the measure being the most appropriate for this project, for the reasons outlined above. This survey format (using different Maths questions) has been developed and validated for post-GCSE students.[26] The survey will be administered online in May 2019.

As mentioned above, this trial will also make use of a delayed post-test to capture any long-term effects of the programme. This will be considered secondary analysis. The following measures will be collected for the second, cohort whose assessment will take place in summer 2020:

1. **KS5 mathematics GCSE resit performance**, as per the UMS score collected in the same fashion as the first year.
2. **KS5 mathematics GCSE pass rates,** as specified above.

We will also conduct a pooled analysis, in which we combine the data from both cohorts and analyse together, controlling for cohort fixed effects. This will be considered secondary analysis. For the purposes of this evaluation, the KS5 mathematics GCSE resit performance collected for first cohort constitutes the primary outcome.

Analysis plan

Summary

Our primary analysis will focus on KS5 mathematics GCSE resit performance in Cohort 1. This outcome variable will be regressed using a least squares linear model. This approach is detailed below. This will be assessed for KS5 students in summer 2019, with GCSEs to be administered by the settings as per routine school procedure for these exams.

Analysis will be conducted on an intention-to-treat (ITT) basis, including all students in classes of teachers participating in the programme. Analyses will be conducted in Stata version 14, using 2-sided significance tests, at the 5% significance level.

The analysis requires a comparable group of students from each of the settings assigned to the control group. Settings will be asked to nominate teachers to participate in the programme during the recruitment process (prior to being informed of their trial arm

[23] For information on the design and validation of this surveying approach see: Pampaka, M., Kleanthous, I., Hutcheson, G. D., & Wake, G. (2011). Measuring mathematics self-efficacy as a learning outcome. Research in Mathematics Education, 13(2), 169-190. The survey itself is available online at: http://www.teleprism.com/surveys.htm. [Last accessed: 30 April 2018].
[24] Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. Psychological review, 84(2), 191.
[25] Gutman, L. M., & Schoon, I. (2013). The impact of non-cognitive skills on outcomes for young people. Education Endowment Foundation, London, 11.
[26] Pampaka, M., Kleanthous, I., Hutcheson, G. D., & Wake, G. (2011). Measuring mathematics self-efficacy as a learning outcome. Research in Mathematics Education, 13(2), 169-190.

allocation). This will allow us to identify the students who would have received in the intervention, had their setting been assigned to the treatment group.

Baseline characteristics

Baseline characteristics observed in the NPD (age, KS2 mathematics scores, KS4 mathematics GCSE grades expressed as effect sizes, Ever 6 FSM) will be summarised by treatment arm.

Trial completion

A CONSORT diagram will be used to present the summary of the flow of eligible students and their settings from recruitment through randomisation, post intervention assessment and analysis. The number of children and schools included or excluded at each stage will be clearly stated and the reasons for exclusion will also be stated.

Primary analysis

Analysis will be carried out using an OLS regression,

$$Y_{is} = \beta_0 + \beta_1 \cdot T_{is} + X_{is} \cdot \alpha + \epsilon_{is}$$

where:

- Y_{is} is the outcome for the KS5 mathematics GCSE resit performance for individual i, in setting s, measured by UMS score;
- T_{is} is a binary indicator for the treatment for individual i, in setting s (1 if the student is treated and 0 if not);
- X_{is} is a vector of individual-level and setting-level covariates including a dummy variable for the type of setting (FE College, 6th Form College, School), the baseline attainment measured through separate KS2 raw mathematics scores and KS4 mathematics GCSE grade; and
- ε_{is} is the cluster-robust error term, for individual i in settings s, clustered at the setting level (assuming the errors are correlated within setting and reflecting the design of the study).

While UMS scores are bounded between 0 and 100, we assume that the response to the treatment will be locally linear so an OLS will be appropriate (in any case OLS gives the best linear approximation). The primary analysis will convert effect sizes into Hedges' g using the following transformation:

$$\text{Hedges' } g = \frac{M_1 - M_2}{SD^*_{pooled}}$$

Where M_1 - M_2 is the difference in mean UMS scores between the treatment and control group, and SD^*_{pooled} is the pooled and weighted standard deviation.

Analysis will be ITT, in which we test the hypothesis that being assigned a place on the programme has a beneficial impact on attainment.

Secondary analysis

For the secondary analysis, we will replace Y_i in our specification of the primary analyses with variously:

- Cohort 1 GCSE mathematics pass or fail binary measure (1 if grade 4 or above, 0 if not)
- Cohort 1 student self-efficacy score
- Cohort 2 KS5 mathematics GCSE resit performance, measured by UMS score
- Cohort 2 GCSE mathematics pass or fail binary measure (1 if grade 4 or above, 0 if not)

If students from Cohort 1 fail their KS5 GCSE maths examination, they may appear in our Cohort 2 sample given students must continue resitting until they pass, or turn 19. Based on previous data, this could be up to 70% of students given that only roughly 30% of resit students pass in any given academic year. We will exclude all of these students from our Cohort 2 analysis. The rationale being that in the event that the intervention has an impact on student pass-rates, the follow-through of students from Cohort 1 to Cohort 2 will be differential between our treatment and control groups. Consequently, the follow-through would be correlated with our treatment and confound our results. Therefore, it would only be acceptable to include these students if the intervention has no effect, and this would require setting an arbitrary threshold for what "no effect" means in this scenario.

We will also conduct the analysis for the subgroup of pupils who have been registered for free school meals at any point in the last six years (Ever 6 FSM) in the NPD (using the *EVERFSM_6_P* variable), using the same model as our primary analysis, with the addition of an interaction between treatment assignment and FSM status, to assess whether there is a significant difference in the treatment effect between FSM students and non-FSM students. It must be noted that for any mature learners (aged 19+) in our sample, this measure is not defined (ie. is always set to 0). Therefore, we will exclude mature learners (who are identifiable as we are collecting Date of Birth for each student) from this subgroup analysis. A further subgroup analysis will be completed for type of setting, to identify any differences between pupils retaking GCSE exams in colleges vs 6th forms.

Pooled analysis

As mentioned above, we will conduct a pooled analysis in which we combine the data from both cohorts and analyse together. We will analyse using the OLS regression specified below:

$$Y_{isc} = \beta_0 + \beta_1 \cdot T_{isc} + \beta_2 \cdot C_{is} + X_{is} \cdot \alpha + \epsilon_{isc}$$

where:

- Y_{isc} is the outcome for the KS5 mathematics GCSE resit performance for individual i, in setting s and cohort c, measured by UMS score;
- T_{isc} is a binary indicator for the treatment for individual i, in setting s and cohort c (1 if the student is treated and 0 if not);
- C_{is} is a binary indicator of the cohort for individual i, in setting s (1 if the student is in cohort 2, and 0 if not);
- X_{is} is a vector of individual-level and setting-level covariates including a dummy variable for the type of setting (FE College, 6th Form College, School, Training Provider), the baseline attainment measured through separate KS2 raw mathematics scores and KS4 mathematics GCSE grade; and
- ε_{isc} is the cluster-robust error term, for individual i in settings s and cohort c, clustered at the setting level (assuming the errors are correlated within setting and reflecting the design of the study).

Robustness Checks

We will conduct robustness checks for all primary, secondary and pooled analysis using a Hierarchical Linear Model (HLM) which is an augmented OLS specification that takes into account the hierarchical structure of the data. In this setting, our outcome data is "nested" in that students are clustered in an organised manner (classes within settings).

Treatment effects in the presence of non-compliance

We will use an instrumental variables approach for a Complier Average Causal Effect (CACE) analysis, which will provide the average treatment effect for compliers. In the context of the trial, we have defined compliance as the number of Maths-for-Life lessons a participating teacher delivers to students in each of the classes they teach, which can range between 0 and 5 for any given class. In order to estimate the treatment effect for compliers, we will employ an instrumental variables approach. This approach relies on having a variable which is associated with whether someone receives the treatment but not with the outcome variable of interest. This is known as the instrument. In the context of the trial, the instrument is treatment assignment, which is assumed to influence whether you participate in the programme but not the outcome variable in its own right. Two key assumptions need to hold for this approach:

1. Being assigned to the treatment increases participation in the treatment. In this instance, teachers may only participate in the programme if they are assigned to treatment. We believe we will have sufficient participation among treatment group teachers for this assumption to hold. There is no ability to participate in the programme outside of the assignment to the treatment group.
2. Random assignment does not itself impact outcomes. We have no reason to believe that the offer of the programme would improve attainment on its own, but instead believe the impact on attainment is achieved through participation in the programme.

It is important to note that our measure of compliance is both:

(1) Self-reported[27] and therefore may not accurately reflect teachers' fidelity to the programme and,
(2) Does not reflect whether a student received the treatment as, under the assumption the Maths-for-Life lessons were delivered, we do not observe whether any given student has attended that class.

However, this is the only feasible compliance indicator that has been identified. This analysis is likely to generate treatment effects that exceed those generated by ITT (unless the treatment is detrimental).

The CACE estimation would proceed using a standard two-stage least squares (2SLS) approach. We estimate:

$$T_{is} = \gamma_o + \gamma_1 Z_{is} + X_{is}\delta + u_{is} \quad \textbf{(1)}$$

[27] The data will be collected by Lead Teachers from Class Teachers at cluster training days. Class Teachers will be asked to complete register that logs whether or not they have delivered the lesson for that period for any given class that they teach. Therefore, if a teacher teaches multiple classes, multiple flags will be collected for that lesson. This data will be used by the University of Nottingham to monitor delivery, and sent to BIT at the end of the project for use in the CACE analysis.

$$Y_{is} = \beta_o + \beta_1 \widehat{T}_{is} + X_{is}\alpha + \varepsilon_{is} \quad \textbf{(2)}$$

where:

- Z_{is} is a binary indicator for the treatment (1 if the student is treated and 0 if not);
- T_{is} is a continuous variable indicating the number of Maths-for-Life lessons a participating teacher delivered;
- X_{is} is a vector of individual-level and setting-level covariates including a dummy variable for the type of setting (FE College, 6th Form College, School, Training Provider), the baseline attainment measured through separate KS2 raw mathematics scores and KS4 mathematics GCSE grade;
- u_{is} are Newey-West robust standard errors;
- ε_{is} are Baum–Schaffer–Stillman 2SLS errors;
- $\widehat{T}_{is}$ are the predicted levels of compliance with the programme from (1); and
- Y_{is} is the outcome for the KS5 mathematics GCSE resit performance for individual i, in setting t, measured by UMS score.

Implementation and process evaluation methodology

Introduction

This section details the research questions, methods and sampling strategy that will be adopted for the implementation and process evaluation (IPE). The research questions have been developed and prioritised, guided by the pilot evaluation findings and the outputs of the IDEA workshop. The pilot evaluation also allowed us to test a range of methods for their feasibility during this trial. Based on this, we have chosen a sequential mixed-methods approach, combining quantitative administrative and survey data with a set of qualitative case studies.

Rationale

As the intervention comprises several stages of information dissemination - from the Lead Teachers, to the Class Teachers, to their students, it will be important to understand to what extent the Maths-for-Life model is transmitted to pupils as intended. Therefore, fidelity and dosage of the programme will be particularly important to monitor, to understand to what extent teachers and students are exposed to the intervention, challenges in receiving the full dose of the intervention and the effects that this has. The pilot evaluation suggested that we should expect substantial variation in Class Teacher and Lead Teacher knowledge and skills, so monitoring the quality of the delivery will also be essential, to understand how the teacher's ability affects delivery of the intervention, and how it is received by students. Given that Maths resits are a particular focus in further education pedagogy, understanding how the intervention sits in comparison to current pedagogic practices, and other college or school initiatives, will be essential. Finally, interrogating the logic model, in particular the theorised causal mechanisms, will help us gain a more in-depth understanding of what particular active ingredients may be bringing about changes - this is essential to understand for future programme iterations and policy development in this space.

The IPE has been designed to complement the impact evaluation and to focus on the causal mechanisms and some key moderating factors identified the logic model. Administrative data will be used to assess dosage at two levels: the amount of training received by Class Teachers and the amount of lessons received by students. The latter will be used in the CACE analysis of the impact evaluation, as described above. Surveys with Lead Teachers will give information on fidelity of delivery and the responsiveness of Class Teachers to the

intervention. This will help us to understand the extent to which the estimated effect from the impact evaluation is a result of the programme as it is intended to be delivered and received. Survey data from Class Teachers on programme differentiation will support this analysis; for example, if we find that teachers in the control group have engaged in intervention-like activities, this may suggest that the estimated effect of the intervention is an underestimate. Qualitative case studies will be combined with a more detailed survey of Class Teachers to explore quality of delivery and causal mechanisms. By better understanding the variation in quality of delivery, we will again be able to better interpret the estimated effect and will also be able to provide insights for programme development. Observations of the intervention and interviews with a purposive sample of key stakeholders will allow us to interrogate the causal mechanisms, furthering our understanding of the way that intervention achieves its effect (if any effect is observed).

Research questions

The process study will focus on six core questions relating to factors and dimensions affecting implementation.

1. **Fidelity:** To what extent do implementers adhere to the intended model? In particular:
    - 1.1. To what extent do Lead Teachers adhere to the PD programme?
    - 1.2. To what extent do Class Teachers adhere to the lesson plans?
    - 1.3. What are the barriers to and facilitators of adherence?
    - 1.4. How does non-adherence/adaptation seem to influence outcomes?[28]

2. **Dosage:** How much of the intervention is delivered and received? In particular:
    - 2.1. To what extent do Class Teachers receive the recommended amount of PD?
    - 2.2. To what extent do students receive the recommended amount of Maths-for-Life lessons?
    - 2.3. What factors contribute to any variation in session number and length?

3. **Responsiveness:** To what extent do participants engage with the intervention? In particular:
    - 3.1. To what extent do Class Teachers engage in PD activities?
    - 3.2. To what extent and how is Class Teachers' general practice[29] altered by the programme?
    - 3.3. To what extent do students engage in Maths-for-Life lessons?
    - 3.4. Are there sufficient resources and support for Class Teachers (e.g. extension materials for more able students and use of Teaching Assistants) to allow for effective differentiation in lessons?

4. **Programme differentiation:** To what extent is the intervention distinguishable from existing practice? In particular:
    - 4.1. Have Class Teachers (both intervention and control) received PD of a similar nature (either in the past or during the intervention period)?
    - 4.2. Is the Maths-for-Life teaching approach significantly different from Class Teachers' current practice? If so, how?
    - 4.3. Do control group teachers receive PD of a similar nature?

[28] See TIDieR framework (Appendix 1) for hypothesised adaptations.
[29] I.e. Outside of M4L lessons.

5. **Quality:** How well is the intervention delivered?
    5.1. Can Lead Teachers effectively facilitate Cluster PD sessions?
    5.2. Are the five key pedagogies used effectively by Class Teachers in the delivery of Maths-for-Life lessons?
    5.3. What factors contribute to variation in implementation quality?

6. **Causal mechanisms:** Are the hypothesised mediating mechanisms present? In particular:
    6.1. Are the hypothesised mechanisms that arise from the PD present?
    6.2. Are the hypothesised mechanisms that arise from the Maths-for-Life lessons present?
    6.3. Are there alternative or complementary mechanisms at play?

Methods

Data will be collected using the methods described below. Qualitative case studies will be carried out with a small sample of schools and colleges in the intervention group (see following section for sampling strategy). The methods will be applied sequentially, to allow findings from each stage to inform the approach to the subsequent stage. In particular, themes that emerge from the administrative data and the online surveys will be used to support the sampling and topics of focus in the case studies.

The table below summarises the outcome measures and methods that will be used to answer the research questions.

Research Question	Outcome measure	Method
1. Fidelity: To what extent do implementers adhere to the intended model?		
1.1 To what extent do Lead Teachers adhere to the PD programme?	Lead Teacher adherence during PD sessions - level of adherence to session plan	PD session observations
1.2 To what extent do Class Teachers adhere to the lesson plans?	Class Teacher adherence during lessons - level of adherence to lesson plan Lead Teacher perception of Class Teacher adherence	Lesson observations Lead Teacher interviews and survey
1.3 What are the barriers to and facilitators of adherence?	Lead Teacher behaviour during PD sessions Class Teacher behaviour during lessons Lead Teacher perceptions of barriers and facilitators Class Teacher perceptions of barriers and facilitators	PD session observations Lesson observations Lead Teacher interviews Class Teacher interviews
1.4 How does non-adherence/adaptation seem to influence outcomes?[30]	Level of understanding demonstrated by Class Teachers during periods of non-adherence/adaptation	PD session observations

[30] See TIDieR framework (Appendix 1) for hypothesised adaptations.

	Level of understanding demonstrated by students during lessons during periods of non-adherence/adaptation Lead Teacher justifications for non-adherence/adaptation Class Teacher justifications for non-adherence/adaptation	Lesson observations Lead Teacher interviews Class Teacher interviews
2. Dosage: How much of the intervention is delivered and received?		
2.1 To what extent do Class Teachers receive the recommended amount of PD?	Number of sessions delivered by each Lead Teacher Length of each session delivered Class Teacher attendance at each session	Project Lead session log
2.2 To what extent do students receive the recommended amount of Maths-for-Life lessons?	Number of sessions delivered by each Class Teacher to each of their Maths-for-Life classes[31] Length of each lesson delivered in each setting to each of their Maths-for-Life classes	Lead Teacher lesson log
2.3 What factors contribute to any variation in session number and length?	Lead Teacher perceptions of contributing factors Class Teacher perceptions of contributing factors	Lead Teacher interviews Class Teacher interviews
3. Responsiveness: To what extent do participants engage with the intervention?		
3.1 To what extent do Class Teachers engage in PD activities?	Class Teacher observed engagement during PD sessions Lead Teacher perceptions of Class Teacher engagement Class Teacher perceptions of personal and peer engagement in sessions	PD session observations Lead Teacher interviews and survey Class Teacher interviews and survey
3.2 To what extent and how is Class Teachers' general practice[32] altered by the programme?	Student perceptions of their teacher's general practice Class Teacher perceptions of their general practice	Student interviews Class Teacher interviews and survey
3.3 To what extent do students engage in Maths-for-Life lessons?	Student observed engagement during M4L lessons Class Teacher perceptions of student engagement	Lesson observations Class Teacher interviews and survey

[31] This will be used for the compliance indicator in the analysis of treatment effects in the presence of non-compliance.
[32] I.e. Outside of Maths-for-Life lessons.

3.4 Are there sufficient resources for Class Teachers (e.g. extension materials for more able students) to allow for effective differentiation in lessons?	Student behaviour during M4L lessons - do all students have sufficient work to occupy them for the lesson? Class Teacher perceptions of resources	Lesson observations Class Teacher interviews and survey
4. Programme differentiation: To what extent is the intervention distinguishable from existing practice?		
4.1 Have Class Teachers (both intervention and control) received PD of a similar nature (either in the past or during the intervention period)?Do Class Teachers receive PD of a similar nature?	Class Teacher self-reported engagement in other PD	Class Teacher interviews and survey
4.2 Is the Maths-for-Life teaching approach significantly different from Class Teachers' current practice? If so, how?	Class Teacher perception of personal practice	Class Teacher interviews and survey
4.3 Do control group teachers receive PD of a similar nature?	Control Teacher self-reported PD participation	Control Teacher Survey
5. Quality: How well is the intervention delivered?		
5.1 Can Lead Teachers effectively facilitate Cluster PD sessions?	Lead Teacher behaviour - adherence to facilitator framework Class Teacher perceptions of PD quality	PD session observations Class Teacher interviews and survey
5.2 Are the five key pedagogies used effectively by Class Teachers in the delivery of Maths-for-Life lessons?	Class Teacher behaviour - use of five pedagogies Student perception of lesson quality	Lesson observations Student interviews
5.3 What factors contribute to variation in implementation quality?	Class Teacher perceptions of contributing factors Student perceptions of contributing factors	Class Teacher interviews Student interviews
6. Causal mechanisms: Are the hypothesised mediating mechanisms present?		
6.1 Are the hypothesised mechanisms that arise from the PD present?	Class Teacher discussion during PD sessions Class Teacher behaviour during lessons Student discussion and behaviour during lessons Class Teacher perceptions of the nature of the intervention Student perceptions of the nature of the intervention	PD session observations Lesson observations Lesson observations Class Teacher interviews and survey Student interviews

6.2 Are the hypothesised mechanisms that arise from the Maths-for-Life lessons present?	Class Teacher perceptions of the nature of the intervention Student perceptions of the nature of the intervention	Class Teacher interviews and survey Student interviews
6.3 Are there alternative or complementary mechanisms at play?	Class Teacher behaviour during PD sessions Class Teacher behaviour during lessons Student behaviour during lessons Class Teacher perceptions of the nature of the intervention Student perceptions of the nature of the intervention	PD session observations Lesson observations Lesson observations Class Teacher interviews and survey Student interviews

Administrative data: Dosage data will be collected at two levels. The Project Leads from the University of Nottingham will collect a session log that records the number and duration of PD sessions delivered by each Lead Teacher. Class Teacher attendance rates will also be recorded for each session. Lead Teachers will collect a lesson log that records the number and duration of lessons delivered by each Class Teacher to each of their Maths-for-Life classes. The number of lessons delivered by each Class Teacher will be used for the compliance indicator in the CACE analysis (see 'Analysis Plan' above for more details).

Online surveys: A brief quantitative survey will be issued to Lead Teachers at the end of the intervention to provide data on fidelity and the perceived responsiveness of Class Teachers. A more extensive quantitative survey will be issued to Class Teachers at the end of the intervention that covers all research themes except dosage. Intervention and Control Group Teachers will be asked to complete a brief quantitative survey to establish whether or not they have engaged in intervention-like activities during the period of the trial (for Intervention group Class Teachers, these questions will be integrated into their wider survey).[33]

Case studies: To gather in-depth qualitative insights across all themes of the IPE, six case studies will be conducted that combine observations of PD sessions and Maths-for-Life lessons with interviews with Lead Teachers, Class Teachers and students. The unit of case study will be defined by the Class Teacher, and case studies will be situated within two regional clusters.

- Observations

[33] Whilst a student survey would have also been a useful source of data for the IPE, this method has been ruled out due to the already-high surveying burden placed on students for the secondary outcome data collection.

We will conduct selective, systematic observations[34], to ensure we are capturing the key information to help us answer our research questions.[35] Observations will be partly supported by quality frameworks developed by the University of Nottingham. For classroom observations, this will entail using the Maths-for-Life '5 Key Pedagogies' (which describe the desired teaching approaching), alongside the relevant lesson plan to assess quality and fidelity respectively. The University of Nottingham are producing an equivalent quality framework for Lead Teacher PD sessions, that will be used alongside PD session plans to assess quality and fidelity of this part of the model respectively. Observations will also be used to assess the responsiveness of Class Teachers (to the PD) and students (to lessons). Whilst engagement will not always be visible to the observer, we will attempt to partially infer it from indicators such as the level of attention shown during presentations and the level of questioning and on-topic discussion from and amongst participants. We will also triangulate these findings with interview data, and administrative data. Finally, observations will support our evaluation of the hypothesised causal mechanisms in both PD sessions and lessons. To do this we will look for discussion and behaviour that verifies, contradicts or adds to the hypothesised causal mechanisms in the logic model. For example, in the case of PD, do teachers have discussions that (implicitly or explicitly) reveal a deepening of their understanding of student learning? In the case of lessons, do we observe on-topic student dialogue, and students using representations to understand mathematical structures.

The researcher who will carry out these observations (of both classroom and PD sessions) has experience of designing and delivering classes that adopt a dialogic approach in a range of post-16 settings, as well as experience of training professionals to deliver classes of this nature. This researcher also led the pilot evaluation so has developed a good depth of knowledge of this particular intervention.

One mid-point PD session will be observed in each cluster. Schedule permitting, we will observe this same session across all both clusters. One mid-point lesson will be observed for each of the six case-studied Class Teachers. Again, we will try to observe the same lesson being taught in each case to support comparability between cases.

- Interviews

Lead Teacher, Class Teacher and student perceptions of fidelity, quality, responsiveness and causal mechanisms will also be sought through interviews to build a more complete picture. The two Lead Teachers and six Class Teachers who are observed will also be interviewed for approximately 45 minutes each, after their observation. Approximately five students will be interviewed, one-to-one, from each case study setting (giving a total of 30 students interviewed). Student focus groups were trialled in the pilot evaluation but, due to the low levels of confidence of some students, it was difficult to engage participants equally. As a result, one-to-one interviews have been selected for this evaluation. For all interviews, semi-structured guides will be used to address the relevant research questions (see table above for more details on which research questions will be covered by each set of interviews).

Sampling strategy

All Lead Teachers, Class Teachers and Control Teachers will be asked to complete the relevant online survey. Six Class Teachers will be selected to define the six units for case

[34] Selective observation, in which the researcher focuses on different types of activities to help delineate the differences in those activities (ANGROSINO & dePEREZ, 2000, p.677)
[35] Kawulich, B. B. (2005, May). Participant observation as a data collection method. In Forum Qualitative Sozialforschung/Forum: Qualitative Social Research (Vol. 6, No. 2).

studying. The pilot evaluation and IDEA workshop identified the possibility of substantial variation in the levels of engagement and ability of Lead and Class Teachers to deliver the intervention, which is likely to affect the outcomes. This trial will also introduce new and substantial variation in setting types, which is also hypothesised to be an important moderating factor. Case study pairs will therefore be purposively sampled to aim for variation on the following dimensions:

- Perceived Lead Teacher ability (according to Project Leads)
- Class Teacher engagement (as defined by PD attendance and lesson delivery)
- Setting type (FE College, 6th Form College, School, Training Provider)

This will yield the following sample.

Sampling unit	Estimated sample
Case Studies	
Lead Teacher	2
Class Teacher (Intervention)	6
Setting	6
M4L Class	6
Student	30
Endpoint Survey	
Lead Teacher	All
Class Teacher (Intervention)	All
Class Teacher (Control)	All

Cost evaluation

An estimate of the cost of the intervention (per pupil, per year) will be calculated by the evaluation team. This estimate will focus on cost from the perspective of a participating school or college and will be based on the direct, marginal financial costs of implementing the intervention. This includes anything which the school/college needs to pay for beyond business as usual costs. Time spent by schools/colleges, such time to arrange supply cover for teachers to attend training, but also to prepare for delivery, will be reported separately from the financial costs.

A cost questionnaire will be created for Class Teachers in consultation with the project team at the University of Nottingham, using our understanding of the programme from the pilot year. This questionnaire will be conducted through structured interviews with a sample of six Class Teachers (three from FE Colleges and three in total from: 6th Form College, Schools, and Training Providers). These interviews will be conducted separately to those referred to in the IPE evaluation methodology above. Taking an interview-based approach with a small sample (rather than using an online questionnaire with a larger group) will allow us to probe the level of detail required for an accurate estimate. Adding detailed cost questions to the Class Teacher survey in the IPE has been considered but rejected, as this is likely to reduce the response rate and quality of data collected relating to the other questions.

Ethics and registration

Ethical approval has been sought following Kings College London (KCL) staff ethics approval procedure. It was approved on XXX. BIT has sought KCL ethics approval to ensure the ethics committee is independent of the M4L programme.

Personal data, including teachers' contact details and student demographic and attainment data, will be processed under two separate legal groundings.

Nottingham will seek consent from participating teachers to share their contact details with BIT as part of the registration process.

Settings will share student data with BIT, using legitimate interests as the lawful basis. The argument to justify the use of legitimate interests has three parts to it:

1. There is a legitimate interest: The processing of student data will support the improvement of teaching practices which will benefit future students of Maths GCSE resits.
2. Processing is necessary: BIT can't assess the efficacy of the Maths-for-Life programme without processing the relevant student data.
3. Balance against student interests: Students should reasonably expect this type of processing of their data (schools and colleges regularly use attainment data to support student learning, and aggregated attainment data is regularly used to assess the performance of schools, colleges and the wider UK education system), and the processing of the data will not cause any foreseeable harm to students - particularly as we will not be assessing the performance of individual students.

Students will be given an information sheet during their first maths class, which tells them about the project, the data we'll be collecting and the purposes of data processing. This sheet will give them the opportunity to object to BIT collecting and processing their data. For those who object, we will not collect or process their data. It may be the case that some students miss this information session. In these cases, BIT will still collect the student's data from the school/college, and will ask the student's teacher to give them an information sheet as soon as possible after this to let them know what we have done (and what we intend to do regarding attainment data). At this point, they will again have the opportunity to object and have their data withdrawn from the trial.

The trial will be registered at www.controlled-trials.com.

Data protection

We will ensure that all project related data will be handled in line with our Data Protection and Information and Personal Data Security Policies. We are Cyber Essentials Plus certified and only use secure transfer solutions for sending and receiving data. Once received, all data will be stored and processed securely, with access limited to the staff working on this project to protect data subjects in this trial.

Personnel

Project team

- Geoffrey Wake - University of Nottingham
- Michael Adkins - University of Nottingham
- Matt Woodford - University of Nottingham
- Sheila Evans – University of Nottingham

Evaluation team

- Michael Sanders (Principal Investigator) – BIT
- Pantelis Solomon – BIT
- David Nolan – BIT

- Patrick Taylor – BIT
- Jess Heal – BIT
- Louise Jones - BIT

The teams will have the following roles within the evaluation:

Design of the trial

- sample size calculations – BIT (David Nolan, Pantelis Solomon, Michael Sanders)
- refinement of randomisation approach – BIT (David Nolan, Pantelis Solomon, Michael Sanders)

Delivery of the intervention

- recruitment of colleges and schools – University of Nottingham (Geoffrey Wake, Matt Woodford, Michael Adkins, Sheila Evans)
- delivery of intervention – University of Nottingham (Geoffrey Wake, Matt Woodford, Michael Adkins)

Measurement and collection of outcomes

- KS5 GCSE mathematics UMS Scores – BIT (David Nolan, Louise Jones)
- KS5 GCSE mathematics pass-rates – BIT (David Nolan, Louise Jones)
- self-reported; self-efficacy – BIT (Patrick Taylor, Louise Jones)

Impact analysis – BIT (David Nolan, Pantelis Solomon, Michael Sanders)
Qualitative analysis – BIT (Patrick Taylor, Jessica Heal)

Risks

The key risks to the trial are listed below.

- **Setting attrition after randomisation reduces the integrity of the experimental design**. To reduce the risk of drop-out, it will be important to ensure settings are well-informed about the programme and the trial from the start, so that they are clear as to what is expected of them before they commit to taking part. Schools and colleges will be asked to sign a Memorandum of Understanding (MOU) as a signal of their commitment. Settings will not be informed of their trial arm allocation until after preliminary student data is submitted to BIT. It will also be important to maintain good communications with settings throughout the project in order to maximise retention. There may also be difficulties in recruiting settings to the trial. Records will be kept of settings approached and where possible, their reasons for not participating, to provide an indication of external validity.
- **Failure to collect primary outcome measure if pupils are not present on the day of testing**. This may also reduce the sample size by reducing the number of pupils for whom we are able to obtain a post-test; furthermore, it may introduce some bias if it is a non-random group of pupils who are absent. As academic attainment will be assessed through GCSE exams, the risk of low turnout is much lower for the primary outcome.
- **Failure to collect baseline and primary outcome data from settings.** This risk is especially high in settings allocated to the control group, who may have lower engagement with the evaluation due to their allocation. Previous experience working

in the FE college domain has illustrated the difficulties with collecting data directly from colleges, for a multitude of reasons including staff turnover and administrative challenges within colleges. To ensure BIT can track students from both trial arms in the first cohort, settings will only be informed of their trial arm allocation after sending baseline data to BIT. BIT will also employ additional research staff during the data collection periods to ensure regular contact with settings is maintained, and deadlines for data collection are met. The University of Nottingham has also included provision of specified student data, including UPNs and GCSE UMS scores as part of the MOU, with a financial incentive provided to control group settings to uphold this agreement.

- **There is a possibility that the delivery of the intervention will vary across settings**. Given the delivery channel is through small clusters of teachers delivery is likely to vary. However, this reflects the reality of implementing such a programme; impact estimates therefore relate more to the type of treatment likely to prevail in practice rather than that which might be observed under ideal conditions. Nevertheless, understanding treatment variation is important and will be explored as part of the process evaluation.

Timeline

Date	Activity	Staff responsible/ leading
09 April 2018	IDEA workshop	BIT, EEF, UoN
29 June 2018	Introductory session for all participating teachers	Geoffrey Wake & Matt Woodford
By end July 2018	All settings (FE Colleges, 6th Form Colleges, Schools, Training Providers) are recruited	Geoffrey Wake, Matt Woodford & Sheila Evans
11th October 2018	Settings provide data for all pupils participating in the trial	David Nolan & Louise Jones
12th October 2018	Settings are randomly allocated to two groups by BIT, balance checks are conducted and settings are then informed of their allocation	David Nolan
W/C 15th October 2018	Intervention begins (PD programme), IPE commences	Geoffrey Wake, Matt Woodford, Sheila Evans, Patrick Taylor & Jessica Heal
W/C 13th November 2018	Settings provide updated data for pupils participating in the trial	David Nolan & Louise Jones
Autumn Term 2018	Project Lead session log data collection	Geoffrey Wake, Matt Woodford & Sheila Evans
	Lead Teacher lesson log data collection	Geoffrey Wake, Matt Woodford & Sheila Evans
	Sampling of case studies	Patrick Taylor & Jessica Heal
1st December 2018	BIT collects baseline data from the NPD for Cohort 1	David Nolan and Louise Jones
Spring Term 2019	PD observations	Patrick Taylor & Jessica Heal
	Lesson observations	Patrick Taylor & Jessica Heal
	Interviews	Patrick Taylor & Jessica Heal
	Endpoint online IPE surveys	Patrick Taylor & Jessica Heal
30th May 2019	BIT collects secondary outcome data (Self-efficacy survey)	Patrick Taylor & Jessica Heal
24th June 2019	Intervention ends	Geoffrey Wake, Matt Woodford & Sheila Evans
2nd July 2019	1st cohort students sit GCSEs	N/A
1st September 2019	BIT begins collecting primary outcome data (UMS scores from schools/ colleges) for first cohort.	
31st October 2019	Settings provide data for pupils participating in the trial in the second year	David Nolan and Louise Jones
1st December 2019	BIT collects baseline data from the NPD for students in the second cohort of the trial	David Nolan and Louise Jones

1st December 2019	BIT collects secondary outcome data (student grades) from the NPD for first cohort	David Nolan and Louise Jones
1st December 2019	BIT finalises primary outcome data (UMS scores from schools/ colleges) for first cohort.	David Nolan and Louise Jones
31 March 2020	BIT submit draft report on first cohort	Michael Sanders and David Nolan
2nd July 2020	2nd cohort students sit GCSEs	N/A
31 August 2020	BIT submits final report on first cohort	Michael Sanders and David Nolan
1st September 2020	BIT begins collecting primary outcome data from settings for second cohort	David Nolan and Louise Jones
1st December 2020	BIT collects secondary outcome data (student grades) from the NPD for second cohort	David Nolan
1st December 2020	BIT finalises primary outcome data collection from settings for second cohort	David Nolan and Louise Jones
28th February 2021	BIT submits report appendix on second cohort	Michael Sanders and David Nolan
31st March 2021	BIT submits final appendix on second cohort; BIT submits data to FFT archive and updates the trial register with results	Michael Sanders and David Nolan

Appendices

Appendix 1: Template for Intervention Description and Replication (TIDieR): Maths-for-Life

Why (rationale/theory)

The goal of Maths-for-Life is to improve maths GCSE outcomes for students resitting their exams. All students who fail to gain a grade C (or grade 4) or above in their maths GCSE are required to continue working towards this goal in post-16 education until the age of 18. In 2017, the proportion of students aged 17 and above who achieved a grade C or above in GCSE maths was 24.4%.

Building on the work of Malcolm Swan[36], the Maths-for-Life intervention aims to develop a more student-centred classroom, focussed on problem solving and dialogic teaching. Carefully designed activities create cognitive conflict, expose common mathematical misconceptions and improve dialogue between students and teachers. Pictorial representations are also used support problem solving. This approach is designed to deepen student understanding of mathematical concepts, increase their ability to solve problems and increase their sense of mathematical self-efficacy.

Who (recipients)

There are two groups of intervention recipient:

1. Teachers of maths GCSE resit students, who participate in a programme of professional development (PD) and;
2. Maths GCSE resit students, who participate in Maths-for-Life lessons as part of their resit curriculum. This group includes a small proportion of adult learners.

What (materials)

There are two sets of materials that support delivery of the intervention:

1. Lead Teacher pack: Resources for Lead Teachers to support their cohort of Maths-for-Life teachers.
2. Class Teacher pack: Resources for teachers to support the delivery of Maths-for-Life lessons, including information about the pedagogical principles, five lesson plans, lesson resources (e.g. slide decks and materials for students), clips of videos of Maths-for-Life lessons that demonstrate core concepts, and proformas for peer lesson observations.

What (procedures)

The intervention begins and ends with a launch event for Class Teachers. In between these two meetings, the intervention goes through the following lesson study cycle:

1. Clusters of Class Teachers meet to learn about and plan a Maths-for-Life lesson, supported by their Lead Teacher.

[36] Swan, M. (2006). Learning GCSE mathematics through discussion: what are the effects on students?. *Journal of Further and Higher Education*, *30*(3), 229-241.; Swan, M. (2007). The impact of task-based professional development on teachers' practices and beliefs: A design research study. *Journal of Mathematics Teacher Education*, *10*(4-6), 217-237.; Swan, M., & Swain, J. (2010). The impact of a professional development programme on the practices and beliefs of numeracy teachers. *Journal of further and Higher Education*, *34*(2), 165-177.

2. Class Teachers deliver a Maths-for-Life lesson to their own class
3. Class Teachers meet as a whole cluster to observe a peer from the cluster delivering the same lesson
4. Clusters meet again with their Lead Teacher to reflect on the lesson delivered, and to learn about the next lesson to be delivered.

This cycle is completed five times, with a new lesson being delivered and studied each time.

Who (implementers)

There are four key roles in the delivery of the intervention:

1. Class Teachers deliver the Maths-for-Life lessons.
2. Lead Teachers are maths GCSE resit teachers from other settings that facilitate the PD programme for a cluster of Class Teachers.
3. Project Leads from the University of Nottingham support the Lead Teachers through initial training and ongoing monitoring and troubleshooting of the programme (including some lesson observations).
4. Coordination of the project is supported by an administrator from the University of Nottingham.

How (mode of delivery)

The five lessons are taught as part of the normal GCSE maths resit curriculum. The five PD sessions and peer observations are delivered during Class Teachers' professional development time.

Where (setting)

Lessons are taught in regular classrooms in schools and colleges. Cluster PD sessions take place in a range of regional locations, such as participating schools and colleges.

When and how much (duration and dosage)

The five lessons are delivered between November and April. Each lesson is designed to last for one hour. The PD requires a total of 6 days of teacher time and breaks down as follows:

- Launch event: 0.5 days
- Lesson planning and reflection: 1 day per lesson (5 days in total)
- Closing event: 0.5 days

Tailoring (adaptation)

The following adaptations are expected to take place at the classroom level:

- Lessons are planned to last 1 hour but some may last for up to 2 hours. The time taken to deliver each component of the lesson will also vary.
- Whilst questioning guides are provided for Class Teachers, the exact type and number of questions asked of students will vary substantially.
- Activities are designed for groups of students, but the size of groups will vary from two upwards. In some (rare) cases, students may work on their own.
- Some classes will be supported by Teaching Assistants / Learning Support Assistants.

- The following adaptations are expected to take place at the cluster level:

- The timing of cluster PD days
- The duration of cluster PD days
- The types of location that are used for cluster PD days
- The content of PD day discussions (responding to the needs and experiences of the group)

Three core components of the intervention should not change in any setting:

- The key pedagogical principles
- The lesson study structure of the PD
- The lesson materials provided to students

How well (planned)

Strategies to maximise effective implementation operate at three levels:

1. Class Teachers are given clearly defined lesson plans, supported by an in-depth programme of PD.
2. Lead Teachers monitor the delivery of lessons in their cluster through a lesson log, offering troubleshooting support where necessary.
3. Project Leads provide training, written guidelines and ongoing troubleshooting support to Lead Teachers as necessary.

How well (actual)

Effective implementation is expected to be moderated in particular by variation in:

- The ability of Lead Teachers and their level of belief in the pedagogy
- The ability of Class Teachers

Appendix 2: Memorandum of Understanding

EEF Maths-for-Life

Memorandum of Understanding (MoU)

http://m4l.org.uk/

Name of College/school _____

Date_____

The Maths-for-Life project is based in the Centre for Research in Mathematics Education at the University of Nottingham. It is investigating the efficacy of a professional development programme that aims to improve outcomes in GCSE Mathematics re-sit examinations. The project is funded by the Education Endowment Foundation[37] (EEF) as part of a joint initiative with JP Morgan. EEF has appointed the Behavioural Insights Team[38] (BIT) to evaluate the project.

This document sets out the relationship between the University of Nottingham, the Behavioural Insights Team, teachers and their colleges and schools participating in the Maths-for-Life project between June 2018 and October 2020. All colleges and schools taking part in the project are asked to confirm their agreement to this relationship by signing their copy of this document. Teachers participating and an appropriate senior manager should sign this document to confirm that the college/school will fulfil the requirements of their participation in the project as identified in this document.

By signing this Memorandum of Understanding you agree to allow the University of Nottingham Maths-for-Life team to share your contact data with the project evaluation team, BIT.

Note: See attached a separate Data Sharing Agreement to be signed between your school/college and BIT covering the data sharing requirements described below.

[37] http://www.educationendowmentfoundation.org.uk
[38] http://www.behaviouralinsights.co.uk/

A. All colleges/schools agree to:

At sign up
• Identify one or two members of staff who teach GCSE Mathematics resits to take part in the Maths-for-Life intervention in the period October 2018 – June 2020. (Throughout the project data of all GCSE resit students taught by the nominated teachers will be included).
• Securely provide BIT and the University of Nottingham with required information about the college/school.

• Share the **Student Information data** for participating students (that is all GCSE resit students taught by the nominated member(s) of staff) with BIT via a secure platform by 11th October 2018:
  ◦ Unique Pupil Number
  ◦ Name
  ◦ Date of Birth
  ◦ Free School Meal status (Ever6FSM)

• Nominated teachers will engage in a brief online pre-intervention teacher survey.

Throughout Year 1 (October 2018-September 2019) and Year 2 (October 2019-September 2020)
• Provide BIT and the University of Nottingham with updates to student and teacher information (for example, if any of the original students or teachers have left the college/school or changed classes).

• If selected as a case-study college/school facilitate researchers to:
  ◦ observe all aspects of the 'professional development' days and collaborative teachers' communities;
  ◦ interview teachers and students;
  ◦ observe teaching and learning.
Note that this will include gaining opt-in consent from all teachers and students participating.

• Engage in a post-intervention teacher survey.
• Support the administration of a paper-based[39] self-efficacy survey with all participating students between April and June 2019. (BIT Researchers will attend your school/college for this purpose).

[39] After MOUs were issued, it was decided to switch to online surveying.

B. If allocated to the Maths-for-Life intervention group, colleges/schools agree to:

Year 1 (October 2018-September 2019)
• Ensure participating teachers attend a whole day project briefing in London on 19/10/2018 and take on responsibility for delivering the agreed work.
• Ensure chosen teachers participate in the local cross college/school development programme coordinated by the Maths-for-Life Lead Teacher. This will include a cycle of lesson briefings, observing the research lessons and a post-lesson discussions. These meets will take place over an equivalent of 5 days in addition to the project briefing day.
• Ensure all five Maths-for-Life lessons are taught to all of the participating GCSE resit classes by 12/04/2019.

• Share the **Student Attainment data** of all participating students (that is all GCSE resit students taught by the nominated member(s) of staff) with BIT via a secure platform by September 2019.
  ◦ GCSE Maths Uniform Mark Scores (UMS)
  ◦ GCSE grades achieved

Year 2 (October 2019-September 2020)

• Share the **Student Information data** for participating students (that is all GCSE resit students taught by the nominated member(s) of staff) with BIT via a secure platform by 31st October 2019:
  ◦ Unique Pupil Number
  ◦ Name
  ◦ Date of Birth
  ◦ Free School Meal status (Ever6FSM)

• Share the **Student Attainment data** of all participating students with BIT via a secure platform by September 2020.
  ◦ GCSE Maths Uniform Mark Scores (UMS)
  ◦ GCSE grades achieved
 	
Note that the scores provided in 2020 will be for the new students of the teachers who participated in the Maths-for-Life development programme in 2019-2020 academic year.

• Raise any concerns with the project administrator (University of Nottingham) about changes at the earliest opportunity. If concerns arise, the resolution of such will be based upon the principle of mutual respect and a desire to find a solution speedily and informally.

University of Nottingham agrees to:
• Provide ‘’professional development” and support from the Maths-for-Life Lead Teachers between October 2018 and March 2019 to aid with the delivery of the materials and provide the support required to complete the activities.

- Provide all Maths-for-Life teacher support and lesson materials.

- Ensure that the research has ethical clearance from the research ethics committee of the University of Nottingham and that this is adhered to throughout the project.

- Store all data safely and securely.

- Inform schools/colleges of the results of their random allocation to intervention or control group by 12/10/2018. This will ensure that there has been sufficient time to collect the required data from schools and colleges before inviting them to the project launch on 19/10/2018.

C. If allocated to the Maths-for-Life control group, colleges/schools agree to:

Year 1 (October 2018-September 2019)

- Share the **Student Attainment data** of all participating students with BIT via a secure platform by September 2019.
    - GCSE Maths Uniform Mark Scores (UMS)
    - GCSE grades achieved

Year 2 (October 2019-September 2020)

- Share the **Student Information data** for participating students with BIT via a secure platform by 31st October 2019:
    - Unique Pupil Number
    - Name
    - Date of Birth
    - Free School Meal status (Ever6FSM)

- Share the **Student Attainment data** of all participating students with BIT via a secure platform by September 2020.
    - GCSE Maths Uniform Mark Scores (UMS)
    - GCSE grades achieved

Note that the scores provided in 2020 will be for the new students of the teachers who participate in 2019-2020 academic year.

University of Nottingham agrees to:

- Pay a total of £1000 to a college/school in the control group.

yment	nount	pected Payment Date	Notes
1		ber 2019	eipt of 2018-19 Student Information data as set out in section A.
2		nber 2020	eipt of 2019-20 Student Information and Student Attainment data as set out in section C.

- Provide the support required to complete the activities mentioned above.

- Ensure that the research has ethical clearance from the research ethics committee of the University of Nottingham and that this is adhered to throughout the project.

- Store all data safely and securely.

- Inform schools/colleges of the results of their random allocation to intervention or control group by 12/10/2018.

AGREEMENT TO PARTICIPATE AND WITHDRAWAL OF PARTICIPATION

Participation in the project by your college/school is voluntary.

By completing, signing and returning this form you confirm your understanding of the project and agree to all aspects of taking part in it. Please make sure to ask any questions, by contacting the Project Administrator (University of Nottingham), about the project before signing.

If your school or an individual from your school would like to withdraw from the project they can do so at any point until the final data is collected (October 2020) by contacting the project Director in the first instance (details below):

Geoff Wake

School of Education,

Jubilee Campus,

University of Nottingham,

Nottingham

NG8 1BB

Geoffrey.wake@nottingham.ac.uk

Signatures

College/school Senior Leader (with authority to commit human resources) By signing this form you are committing to supporting the teacher(s) identified below in pursuit of the Maths-for-Life project and supporting administrative staff in supplying the required data.	Name: Position: Signed: Date:
Participating Teacher 1 By signing this form you are committing to taking part fully in the Maths-for-Life project.	Name: Position: Contact e-mail: Telephone number: Date:
Participating Teacher 2 **(if applicable)** By signing this form you are committing to taking part fully in the Maths-for-Life project.	Name: Position: Contact e-mail: Telephone number:

	Date:
University of Nottingham	Name: Position: Signed: Date: