PROJECT TITLE	Young Enterprise: Mathematics in Context
DEVELOPER	Young Enterprise
EVALUATOR	University of Nottingham, UCL Institute of Education
PRINCIPAL INVESTIGATOR	Geoff Wake, Jeremy Hodgen
PROTOCOL AUTHOR(S)	Jeremy Hodgen, Geoff Wake & Michael Adkins
TRIAL DESIGN	Randomised Controlled Trial, two-arm design: intervention vs. business as usual control
AGE RANGE, KEY STAGE	14-16, KS4
NUMBER OF SCHOOLS	130 (state sector)
NUMBER OF TEACHERS	520 (4 per school)
NUMBER OF PUPILS	13000 (approximately 25 pupils per teacher and 100 pupils per school)
PRIMARY OUTCOME	GCSE Mathematics attainment (UMS scores)

Protocol version history

VERSION	DATE	REASON FOR REVISION
1.1	25/05/2018	1. To reflect changes / amendments to randomisation: randomisation was stratified by region (North/South England) and by school-level everFSM (p.15), and randomisation was conducted in two batches (p.14). 2. Clarification of the legal basis for processing data in accordance with the GDPR (p.26) and related risk added (p.28). 3. Additional 'wrap-up' one-day event added to project description (p.6). 4. Trial registration number added (p.26).
1.0	03/04/2017	N/A

Abstract

"Young *Enterprise: Maths in Context"* is an intervention that seeks to improve children's financial capability, and specifically their financial knowledge and understanding, applied numeracy and problem-solving skills. Three measurement tools will be used to measure the outcomes of this pilot: applied numeracy and problem solving will be measured through Maths GCSE attainment; a financial capability instrument developed from the MAS Financial Capability Outcomes Framework will seek to measure financial knowledge and understanding; and a GCSE subscale of financial and problem-solving will provide a further measurement tool to assess relevant outcomes. The intervention programme is designed to address financial capability teaching targeted at Key Stage 4 (children who are aged 14-16). This evaluation will undertake a two-arm randomised controlled trial testing the intervention against a 'business as usual' equivalent. School recruitment will be in the period October 2016 to May 2017 with teacher training and the trial commencing in schools in September 2017. Final publication of results will be in the spring term of 2020. This protocol outlines the evaluation of the project, including its rationale, objectives, and its three elements: evaluation of impact, process and cost. We discuss the data collection and analysis for all three components, a plan for data release to promote independent scrutiny and overall study limitations.

Table of Contents

Introduction

The Education Endowment Foundation has funded the University of Nottingham to evaluate the *Young Enterprise: Maths in Context* intervention for Key Stage 4 (KS4) learners. The evaluation will undertake a two-arm randomised controlled trial testing the intervention against a 'business as usual' equivalent. The trial will be delivered in schools starting in September 2017, with school recruitment from October 2016 to May 2017, school randomisation in June 2017 and teacher training in September 2017, with the final publication of the results in the spring term of 2020.

Together with a separately published statistical analysis plan, this evaluation protocol sets out the context of the intervention in terms of attainment levels in secondary mathematics, the trial objectives, the research design for both the impact evaluation– addressing the experimental design decisions, trial arms, units of randomisation, sample size calculations, outcome measures which test overall achievement in mathematics, achievement in financial mathematics GCSE items and broader financial capability, and software used – as well as the process evaluation. This will outline the assessment of the fidelity of implementation, how the process evaluation will inform any future effectiveness trial, and the investigation of aspects of 'social validity' e.g. the acceptability, feasibility and utility of the approach. Lastly, the protocol will address issues of transparency and study ethics, outline an evaluation timetable, set out the responsibilities of the evaluation team members, and finally highlight any limitations of the study – including areas for future research.

Background and significance

Given the emphasis on financial education in the new National Curriculum in mathematics citizenship, providing guidance to schools on effective ways of teaching financial capability is both important and timely. We note that the terms *financial literacy* and *financial capability* are often used interchangeably as intended outcomes for financial education. The OECD PISA study defines financial literacy as follows: "Financial literacy is knowledge and understanding of financial concepts and risks, and the skills, motivation and confidence to apply such knowledge and understanding in order to make effective decisions across a range of financial contexts, to improve the financial well-being of individuals and society, and to enable participation in economic life" (OECD, 2013, p.144). In contrast, the Money Advice Service (2013; Bagwell et al, 2014) define financial capability as having three broad dimensions –

connection related to exposure and access to appropriate financial products and channels; *mindset* which looks at financial attitudes and motivations as well as more general attitudes and motivations; and finally *ability* which addresses both financial knowledge and understanding, as well as basic skills such as applied numeracy, literacy and problem solving. In this protocol, we will in general refer to financial capability.

Financial capability is an important life-skill and is likely to make a significant difference to young people's life chances, particularly those from the most disadvantaged groups. Indeed, results from the latest UK Financial Capability Survey (Money Advice Service, 2015) indicate that disadvantaged groups tend to face the greatest challenges relating to managing money and making financial decisions.

Financial education is, however, relatively poorly understood. Atkinson's (2008) review of financial education evaluations concluded that "not only has there been relatively little work in the past on financial capability in the UK or other countries, but also that rigorous, credible policy evaluation showing the incremental impact of financial capability work is difficult to find" (p.ii). Nevertheless, some recent evidence is available. O'Prey & Shephard's (2014) meta-analysis of 21 studies of financial education found greater effects on knowledge (0.18) than on attitudes (0.08) or behaviour (0.07). However, a meta-analysis of 19 studies for the World Bank (Miller et al, 2014) found effect sizes on different behaviours ranging from 0.10 to -0.08. In both meta-analyses, the results should be treated with a great deal of caution as they aggregate widely different studies set in both developed and developing countries (not including the UK). Despite these limitations, these meta-analyses do provide some useful information and we use these meta-analyses (cautiously) to inform our effect size estimates discussed later in this protocol.

Financial education presents a number of challenges: First, financial capability is not equivalent to numeracy or attainment in mathematics. Indeed, whilst numeracy is a necessary element, it is not on its own sufficient for the development of financial capability (e.g., Huston, 2010). Hence, whilst the OECD's PISA survey at age 15 internationally finds a relatively high correlation between mathematics and financial literacy (0.88), the survey also found "wide variations in financial literacy performance for any given level of performance in mathematics and reading, meaning that the skills measured by the financial literacy assessment may go beyond or fall short of the ability to use the knowledge that students acquired from [these] subjects" (OECD, 2014, pp.65-6). Second, the use of real-world contexts in teaching,

particularly mathematics teaching, is fraught with difficulties (e.g., Lave, 1988). Improving financial capability at least in part involves the application and use of numeracy in context, yet evidence suggests that teachers find the application and use of mathematics difficult to teach (e.g., Lesh & Zawojewski, 2007) and that context may in some circumstances hinder, rather than enhance, student learning (Kaminski et al, 2008). Third, financial education may be hindered by students' lack of experience and familiarity with financial products and decisions (Bagwell et al, 2014). Teachers themselves often have poor knowledge and understanding of finance issues and financial capability and this is likely to impact on student learning (Atkinson, 2008).

Project Description

The aim of the project is to test and evaluate, within a rigorous and high quality research framework, the impact of financial education training as part of the Young Enterprise: Maths in Context intervention on attainment in GCSE mathematics and on financial knowledge and understanding when it is delivered by trained teachers through a cascade training mechanism in up to 65 schools against a 'business as usual' control group of up to 65 schools. The intervention is based on an adapted form of an earlier project funded by the London Schools Excellence Fund (PFEG, 2015), which was developed by the charity, Young Enterprise. This earlier project involved a small-scale evaluation of the impact on student attainment involving comparison of the intervention group (260 students) to a control group (101 students) who were taught by the same teachers.[1] This intervention group made greater gains on a levelled GCSE-based assessment.

Intervention

"Young Enterprise: Maths in Context" is an intervention that seeks to improve children's financial capability, and specifically their financial knowledge and understanding, applied numeracy and problem solving skills. The intervention will be delivered by Young Enterprise, the developer, and will take place over the school year 2017-18. The intervention will consist of a series of 10-12 lesson plans, each focused on a specific area of mathematics in the context of financial capability and aimed at Year 10 students, together with a one-day external initial training programme, on-going in-school support from a Young Enterprise consultant mentor,

[1] https://www.london.gov.uk/sites/default/files/pfeg_london_lead_teachers_in_financial_mathematics_final_report.pdf

and a final 'wrap-up' day at the end of the summer term 2018. The lessons, which will be finalised during 2016/17, will be designed for delivery to Year 10 students.

Each intervention school will identify a lead teacher. Lead teachers will be expected to model the teaching approach by implementing the lessons and pedagogies introduced in the training in their own lessons *and* more widely within their schools by providing 'cascade' training to at least three other Year 10 mathematics teachers. The intervention programme will be introduced with one day of external training for lead teachers in regional clusters of between 5 and 12. The training will outline the key pedagogical approaches to teaching mathematics in financial contexts, introduce the lesson plans to the lead teachers and outline possible approaches to cascade training and support for other mathematics teachers.

Each regional training day will be led by an experienced Young Enterprise Consultant, drawn from a pool of personnel already used by Young Enterprise. In addition, the Consultants will provide 'bespoke' in-school support to Lead Teachers, using a framework designed to adapt to the support to the perceived and emerging needs of the school's teachers and pupils. A key aspect of the consultant's work with a school will be supporting the lead teacher with their 'cascade' training. Consultants will provide three days equivalent time of mentoring support delivered over up to eight visits.

Schools will be encouraged to continue teaching financial capability in mathematics to the same students in Year 11 in 2018-19. To facilitate this, they will have access to an online forum and additional online materials.[2]

The Young Enterprise: Maths in Context initiative is summarised in the logic model shown in Figure 1.

[2] These resources are freely available to all teachers.

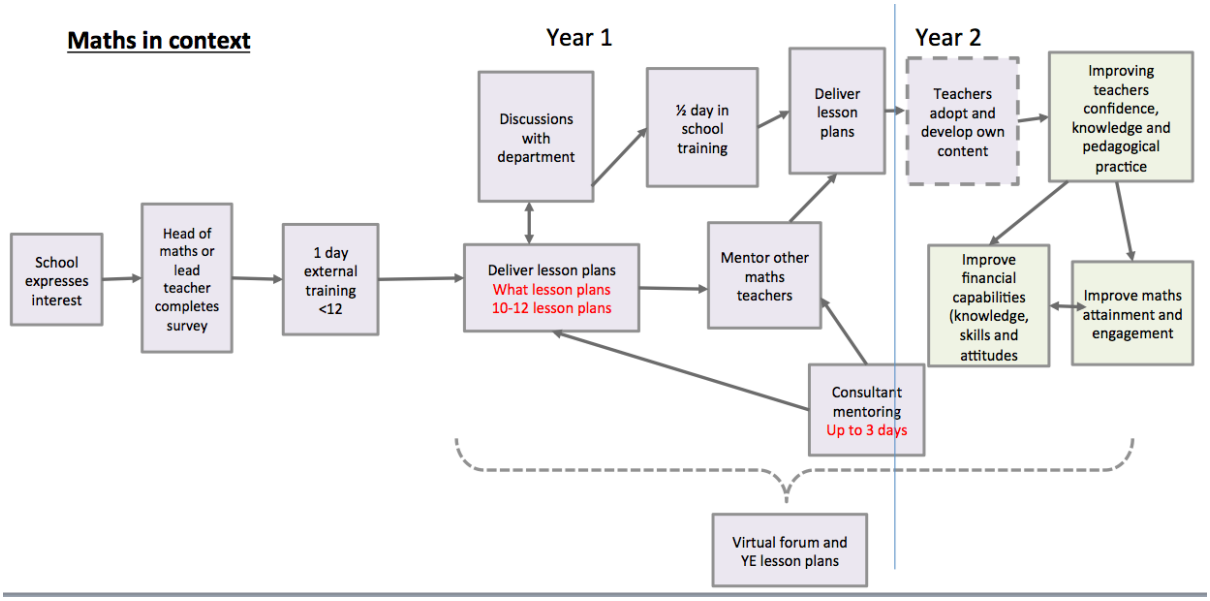


Figure 1: Young Enterprise: Mathematics in Context logic model

Methods

Research Questions

The evaluation will address the following primary research question:

1. Does Young Enterprise: Maths in Context have significant effect on students' attainment in mathematics at GCSE when compared to a business-as-usual control?

In addition, the evaluation will address the following secondary research questions:

2. Does Young Enterprise: Maths in Context have significant effect on students' financial capability when compared to a business-as-usual control?
3. Does the intervention have a significant effect on students' attainment in a GCSE mathematics financial capability sub-scale when compared to a business-as-usual control?
4. Does the intervention have significant effect on students' engagement in mathematics and financial capability when compared to a business-as-usual control?
5. Are the effects on mathematical attainment, financial capability and engagement different for children eligible for free school meals?

6. Are the effects on mathematical attainment, financial capability and engagement different for girls and boys?
7. To what extent are any effects on mathematical attainment, financial capability and engagement mediated by the quantity of mathematics teaching that includes financial contexts and related problem-solving activities?
8. Are there differences in the effects on mathematical attainment, financial capability and mathematical self-efficacy between the lead teachers and other teachers in the intervention schools?

In the process evaluation, we will address the following research questions using an exploratory sequential design [*Italicised comments* in square brackets cross-reference each question to Humphrey et al's (2016) "*Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook*"]:

9. To what extent do the Young Enterprise training sessions, materials and resources enable teachers to overcome students' lack of experience and familiarity with financial products and decisions? To what extent does Young Enterprise enable teachers to use financial contexts to enhance learning of mathematics, including the application and use of mathematics in context? [*Specific aim of the intervention, Fidelity & Adherence, Quality*]
10. To what extent do teachers perceive the intervention resources, training and mentoring to be effective? [*Specific aim of the intervention, Fidelity & Adherence, Quality*]
11. To what extent do intervention schools, lead teachers and other teachers adhere to the intervention in terms of the delivery, the quality of delivery and how much of the intended programme is delivered? To what extent does the intervention enable teachers to use financial contexts and related problem-solving tasks more effectively in mathematics lessons? To what extent is the intervention manageable for teachers to deliver? Is the intervention sufficiently adaptable to the needs of different students and teachers? [*Fidelity & Adherence, Quality, Dosage, Programme differentiation, Reach, Responsiveness, Implementation factors*]
12. How variable is the fidelity and quality of implementation between intervention schools? What school and contextual factors afford or constrain the fidelity quality of implementation? In what ways do schools support the delivery intervention, and

to what extent are the different approaches effective? [*Fidelity & Adherence, Quality, Reach, Programme differentiation, Implementer factors*]

13. To what extent is the intervention scalable? [*Factors relating to implementation*]
14. To what extent does the intervention have an effect on teachers' confidence towards, knowledge of and pedagogical practice relating to financial capability and how it relates to mathematics? To what extent are any effects on mathematical attainment, financial capability and engagement mediated by teachers' confidence towards, knowledge of and pedagogical practice relating to financial capability and how it relates to mathematics? [*All implementation factors*]
15. What does usual practice relating to the teaching of financial capability in mathematics within control schools look like? To what extent (if at all), and how, do teachers in the control schools use financial contexts and related problem-solving tasks in mathematics lessons? [*Monitoring of control group*]
16. To what extent does the intervention encourage teachers to develop effective tasks and strategies to address financial capability in mathematics lessons? [*Specific question relating to the aims of the intervention, Adaptation*]

Design

Development Phase

During the development phase, the evaluation team will contribute to the trialling of lessons and validation of instruments. Each new, and existing, lesson will be trialled on at least three occasions. The evaluation team will:

- Develop a short feedback form for teachers to complete on their experiences of the lesson trials, and
- Observe at least four lesson trials.

Based on these observations and feedback in addition to their curriculum design and pedagogical expertise, the evaluation team will provide feedback to the developers on the new/revised lesson materials, the accompanying guidance to teachers and the suggested intensity / pattern of delivery of the lessons.

Validation of the financial knowledge and understanding instrument

In addition, the evaluation team will develop and validate of a financial knowledge and understanding instrument for this age group of students based on the Money Advice Service's (MAS) Financial Capability Outcomes Framework (Bagwell et al, 2014) in order to assess its feasibility for use as an outcome measure. The MAS framework has been designed to provide a bank of items that can be used to measure financial capability considered broadly and focuses on the three dimensions: *connection, mindset* and *ability*. In the framework, the mathematics element of the ability dimension is measured mainly through the use of a standardised mathematics test of the administrator's choosing. For this reason, we do not think the framework can be used as a sole primary outcome measure, but it has considerable potential to provide a useful secondary outcome measure. The evaluation team will work with the developers and MAS to select an appropriate set of items to pilot as a measure of financial capability for KS4 students. As we understand it, to date the validation of the MAS framework has been restricted to small scale cognitive interviewing and we will, therefore, use Rasch / Item Response Theory modelling to evaluate whether the items form a uni-dimensional scale. The instrument will be piloted twice each to 400 students, and we anticipate removing, adapting and/or adding items between the two administrations. During 2016-17, we will investigate the possibility of administering the instrument online via the Bristol Online Survey or (preferably) using a paper-based machine-readable format. We will review this in Spring 2017 and, if necessary, amend the evaluation protocol at that stage.

Validation of GCSE sub-scale of financial & problem-solving mathematics

During the pilot phase, the evaluation team will examine whether a sub-scale of GCSE items can be created that is sufficiently valid for use as a secondary outcome measure. The necessary data can be collected relatively easily and cheaply from schools during the final data collection phase. This would be a desk research / statistical exercise that would not entail any trialling in school. The validation will involve five stages:

1. Collect 2016 item level GCSE maths score data from 15-20 schools in order to achieve a sample of at least 300 pupils per exam board and per tier,
2. Code the GCSE questions / sub-questions as involving financial and other problem-solving contexts for all GCSE papers,

3. Examine whether the items as a whole (or as either separate financial and problem-solving sub-scales) appear to form (or are reasonably close to forming) a uni-dimensional scale,
4. Investigate ways in which the data can be "standardised" across exam boards and GCSE tiers, and
5. Examine the relationships to GCSE as a whole and to the financial literacy measure (although this would be only on the small sample of schools who pilot this instrument for us).

We expect that standardisation across exam boards on a single tier will be more straightforward because the balance of content and problem-solving appears to be relatively uniform across exam boards (whereas the balance for foundation and higher tier appear to be very different).

Assessing the feasibility of the Maths in Context materials for Post-16 mathematics

Adapting the Maths in Context materials from Key Stage 4 for the post-16 sector presents a number of challenges:

- *GCSE resit pass rates are low and declining:* In 2016, GCSE maths entries from students aged 17 and over rose by 33%, from 131,000 in 2015 to 174,000. However, the pass rate of these students fell from 36% in 2015 to 30% in 2016. Hence, although the absolute number of students achieving a C grade or above increased by just under 10%, this increase is relatively modest in comparison to the scale of the problem overall.

- *Targeting key mathematical topics*: The target group of students have very significant difficulties with key concepts particularly multiplicative reasoning, number and algebra, which underlie aspects of financial capability.

- *Time pressures*: Post-16 GCSE resit mathematics is generally organised differently to school mathematics at Key Stage 4. Teachers typically have a short timeframe in which to cover the entire GCSE curriculum in order to help students improve their GCSE grades. In addition, there are considerable curriculum time and staffing pressures.

- *Limited knowledge and expertise of some post-16 teachers*: The range of knowledge and expertise of post-16 mathematics and numeracy teachers is much wider than in secondary schools. Indeed, Ofsted report variable quality of mathematics provision and teaching

across the FE sector. The materials will need to be adapted to support (and develop) teachers' knowledge and pedagogic skills in both financial issues and mathematics.

- *Student attitudes, motivation and engagement*: The target group of students have already "failed" at school mathematics and many do not feel that mathematics is essential for their career and learning objectives. It is important therefore that these students perceive post-16 mathematics as different and more relevant than school mathematics. The Maths in Context problem-solving approach goes some way towards achieving this.

Because of these factors and their interaction our experience of working with this sector suggests that students do value opportunities to use mathematics effectively in context but this needs to be carefully managed by teachers so that they can focus teaching and learning carefully to ensure engagement with key concepts that will ensure that students are well-prepared to engage successfully with examinations after a relatively short time. Consequently, there needs to be a tight focus by teachers on not only providing insight into the use of mathematics in financial contexts but also using tasks carefully to ensure increased facility with key maths concepts such as multiplicative reasoning and providing students.

The evaluators will work closely with the developers in order to provide feedback that informs the development of materials with particular sensitivity to how they are likely to be implemented in a range of post-16 settings as follows:

1. Provide initial feedback on the revised Maths in Context materials prior to the initial trials including identifying how well the financial context of the materials align with the key mathematical topics (such as proportional reasoning) that this group may struggle with. [Timescale: December-January 2017]

2. Following initial pilot/trials, provide detailed feedback on the basis of an analysis of teacher comments and our own expertise. We envisage that this would involve at least two additional rounds of feedback to Young Enterprise. [Timescale: February-March/April 2017]

3. Observe a small number of lesson trials in order to understand issues of pedagogy and how these can be supported by the design of resources. Our experience of materials design indicates that fidelity in terms of implementation is often difficult to achieve. Lesson trial observation at this stage would be used to inform how the design of the materials might

better support teachers in ensuring essential key didactical principles are implemented effectively. [Timescale: February-March 2017]

4. Produce report on the feasibility and promise of the Maths in Context intervention for Post-16 mathematics, together with brief recommendations about the feasibility of similar interventions at Post-16. This would include an outline of any additional resources, materials and capacity that would be required for effective implementation in either an efficacy or an effectiveness trial. [Timescale: May 2017]

Design of the Main Trial

The intervention has been designed to be a whole school randomised controlled trial for approximately 13,000 pupils in 130 schools. Our aim is to recruit 25 children per class, 4 teachers per school, with school recruitment focused on six regions to be decided during the recruitment process as a result of the patterns of initial expressions of interest.

This experiment will utilise a two armed trial: The Young Enterprise: Maths in Context intervention and a business as usual active control. (See Figure 2.)

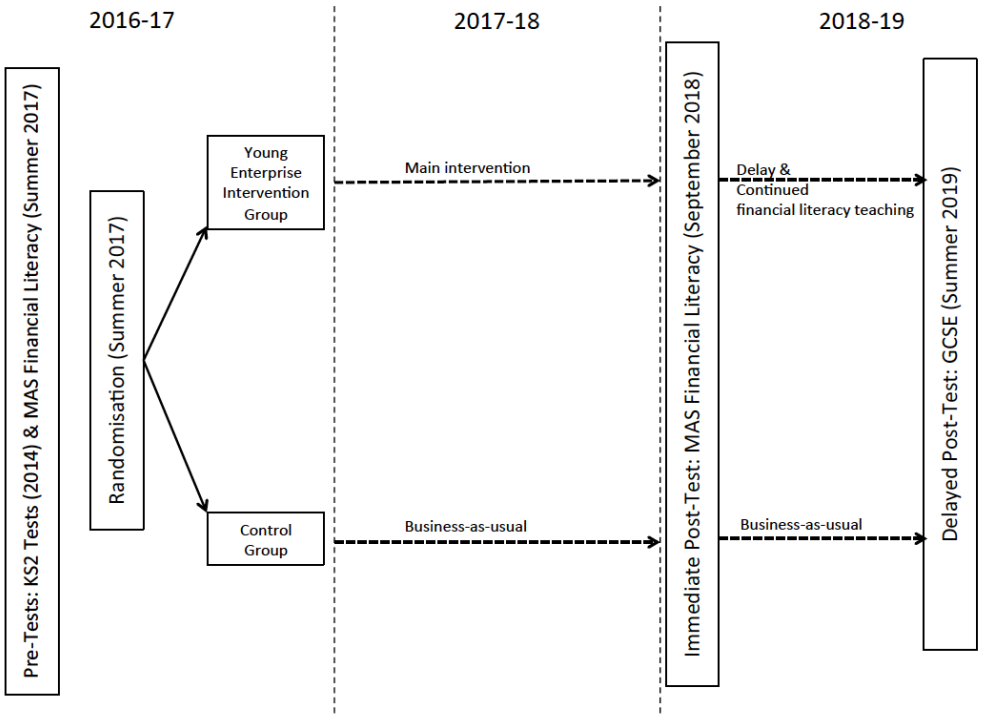


Figure 2: Young Enterprise: Mathematics in Context evaluation design (Note: In the event, due to recruitment difficulties, randomisation was conducted in two batches: 122 schools on 12th July 2017, and 3 on 18th September 2017.)

All schools enrolled in the trial will identify a lead teacher, at least three teachers that will receive cascade training from the trained teacher, and at least four classes of participants. These pupils will sit a pre-test of the proposed financial knowledge and understanding instrument prior to school randomisation. All teachers will provide lessons to all pupils participating for at least one term starting in the autumn term of Year 10). All participants will then have a delayed post-test (GCSE) at the summer of 2019.

Randomisation

We will use a school-level randomisation approach using a block design stratified by geographical area and either everFSM or attainment. We will model this process prior to randomisation and, if necessary amend the protocol in May 2017. We will identify lead teachers prior to randomisation. This will allow an investigation of the cascade training approach will provide much needed data on the potential cost of scaling up the intervention in the future. In addition, the likelihood of contamination in using class based randomisation is very high – from involuntary cascade training on the part of participating teachers, but also between students through sharing material during teaching terms or during the revision period. During 2016-17, we will investigate whether further stratification by school factors (e.g., FSM, GCSE, examination board) is necessary to achieve sufficiently balanced intervention and groups. As a result of this modelling, school randomisation was stratified by region (North/South England) and by school-level everFSM. Due to recruitment difficulties, randomisation was conducted in two batches: 122 schools on 12th July 2017, and 3 on 18th September 2017. The second batch was by simple randomisation.

Participants

All pupils from participating schools enrolled on the GCSE mathematics course at age 14 will be eligible to participate (from Year 10). We will recruit pupils and schools for both foundation and higher tier papers, and will aim to recruit schools using examination scripts from all three major exam boards – Assessment and Qualifications Alliance (AQA), Pearson Edexcel and Oxford Cambridge Recognition (OCR).

Eligibility Criteria

All state schools are eligible as long as the school has not already taken part in Young Enterprise's previous Maths in Context trial, funded by London Schools Excellence Fund

pilot (PFEG, 2015) and can provide a minimum of four classes of year 10s who are eligible for the intervention. Recruitment will aim to maximise the number of schools with an above average proportion of students qualifying as everFSM in order that the proportion of everFSM in the sample as a whole is at least 29.3%.

In addition, in order to be entered into the randomisation the schools will have to have provided:

- Signed Memorandum of Understanding
- Confirmation that consent forms have been sent out and any opt-outs
- Provision of pupil data for those identified as eligible: Class teacher ID, Unique Pupil Number (UPN), Forename, Surname, Date of Birth and Gender
- Pre-test data for all eligible pupils (financial capabilities assessment only)

Incentives

All schools in the control group will be provided with a payment of £1000 on receipt of final GCSE data in Autumn 2019.

Impact Evaluation

The impact evaluation investigates the quantitative evidence for whether there is a consistent positive difference in the post-test scores between participants who have received the intervention as part of the treatment group and those that were recipients of the control condition, conditional on their pre-treatment test score. This subsection will discuss outcome measures, sample size calculations, the main and secondary outcome analyses along with additional sub-group investigations. This will be supplemented by a statistical analysis plan that will contain details on model specifications, estimation procedures and software to be used.

Outcome Measures

Primary Outcomes

We will use GCSE mathematics as the primary outcome measure. In order to improve discrimination and strengthen the statistical modelling, we will use GCSE Uniform Marking Scale (UMS) scores rather than grades and these will collected from directly from schools.[3]

[3] GCSE grades are not designed to form a linear scale. In order to simplify the modelling, we will use the UMS score, which can be modelled using linear regression techniques. The UMS score is a tool that all exam boards use to standardise marks awarded on papers across the different exam boards and paper tiers. The conversions are provided by the exam boards on their websites. For example: http://www.ocr.org.uk/i-want-to/convert-raw-marks-to-ums/

We will use KS2 national test scores in mathematics (KS2_MATPOINTS**)** as a pre-test score for pupils, which will be matched to the UMS score through an extract of the National Pupil Database.

Secondary Outcomes

We propose using two secondary outcome measures subject to the validation process referred to above:

I. An amalgamated scale for financial and problem-solving items from the GCSE mathematics papers. We will collect item-by-item data directly from schools.
II. A bespoke financial knowledge and understanding instrument based on the MAS Financial Capability Outcomes Framework (Bagwell et al, 2014). This secondary measure will be administered by schools in May / June 2016 prior to randomisation, then as a post-test in September 2018. As a secondary measure, we do not consider independent administration (or blinding) to be necessary.

Sample size calculations

We performed two sets of power calculations for the primary and secondary outcomes. We used Raudenbush et al.'s (2011) Optimal Design software to estimate statistical power on the basis of recruiting schools in 2-arm and a 3-level cluster randomised trial with the intervention at level 3 (i.e. the school level). The structure of the intervention is made up of 3 levels – students are clustered in class teachers who are further clustered in schools.

As the research design involves cascade training to multiple teachers, and the likely number of students involved is very high, the greatest change in power is from adding additional schools and so we varied the number of schools that would be part of the intervention.

We fixed the following parameters: α=0.05 (which refers to the probability of rejecting the hypothesis tested when it is true – 5%), 25 students per class and 4 classes per school, intra-cluster correlation for level 2 (class teachers) =0.05 and for level 3 (schools) =0.165 (which refers to the variance between participants with the same TA and for those in the same school). We also included an additional pre-test (KS2 Mathematics score) covariate as a school level aggregate with the assumption that the post and pre-test have a correlation of 0.7 setting the level 3 variance explained at 0.70^2=0.49. This has the effect of reducing the overall variance and boosting the expected statistical power of the study.

This produces an estimated MDES of 0.167 (for 130 schools) and 0.175 (for 120 schools). For the FSM sub-group analysis, a conservative estimate of 8 FSM students per class (approximately 30%) produces an MDES of 0.18.

For testing secondary outcomes, if burden of testing needs to be reduced, testing half the students per class (with a pre-test and assuming a similar post-pre-test correlation) in each school would only result in a very small increase in the MDES. For 130 schools this would be 0.177, and for 120 schools= 0.180. Hence, we propose to survey a minimum of 55 students per school to allow for attrition and to reduce the burden of testing on schools. However, since we anticipate marking the tests using machine readable technology, the impact on the overall costs of testing is minimal, we will review this decision in 2017 and revise the protocol if required.

Analysis

Our analysis will proceed on the basis of descriptives to explore and visualise patterns within the data before building complexity through the use of multilevel modelling to adjust our inferences to take account of the clustering at the teacher and school level. Our proposed analyses are summarised below and will be detailed separately in the statistical analysis plan which will address a number of underlying statistical decisions.

Main outcome

As discussed above, pupil performance in maths will be measured using the UMS scores from GCSE mathematics post-test, with KS2 mathematics scores used as the pre-test. Should attrition become an issue, then we will resort to using the GCSE mathematics grade as recorded in the DfE's National Pupil Database. We will specify an acceptable level of attrition on the basis of simulation of the dataset and amend the protocol in May 2017. Results will be modelled on the basis of intention to treat (ITT). Our analysis will proceed on the basis of starting with simple single-level models incorporating the treatment condition and gradually build complexity by incorporating the pre-test, and adding the varying intercepts for teachers and schools. While fitting a 2-level model can possibly be justified in terms of a more parsimonious model, ignoring the variation at the teacher level would potentially have the effect of underestimating the standard errors of coefficients and would, in our opinion, be an error despite the potential to reduce the complexity of the analysis. It also provides for the possibility of formally testing whether lead teacher class level averages are higher than those receiving the cascade training. As mentioned above, if the variation at the teacher level is small, then we

will consider reverting to a more traditional 2-level model of pupils nested within schools. While the cut-off point is a value judgement, we would suggest that the additional complexity makes little sense where the variance at the teacher level is less than 4% of the unexplained variance in a variance components model.

Effects sizes will be calculated using Hedges *g*.

Subgroup analysis

Additional models of greater complexity will be fitted which will include sex of participant, 'FSM ever' entitlement (defined as any pupil who has ever been classified as in receipt of free school meals), foundation or higher tier paper, as well as interactions between the two original data level variables of treatment and pre-test and the additional sub-group variables. Finally, we will add appropriate group-level predictors including whether the teacher was a school-lead or cascade trained and students' GCSE examination tier. We will conduct an on-treatment analysis based on fidelity. In order to do this, the evaluation and developer teams will jointly develop a measure of fidelity.

Secondary outcomes

We will model the amalgamated GCSE financial capability scale as a separate model (rather than through a multivariate multilevel model) given that it consists of a significant number of questions from the GCSE mathematics paper. Like with the overall score, we will model on the basis of intention to treat, with KS2 mathematics score included as a pre-test.

With the financial knowledge and understanding instrument, we will model this data use a multivariate multilevel model (which allows for multiple outcome variables to be tested simultaneously) using the primary outcome (GCSE score) and the financial capability scale as our second dependent variable. In this model, there are four levels of nesting. Responses to the multiple outcomes are set at the first level providing the structure of the multivariate model, with level two being clusters of individuals, level three being clusters of classes and level four being clusters of schools. In a nutshell, responses are clustered in individuals, who are clustered in classes, who are themselves clustered in schools[4].

Despite the increased complexity of the approach, it offers four significant advantages. Importantly, this model allows for modelling correlations between dependent variables; the standard errors of specific effects tend to be smaller; it allows for the direct comparison of

[4] Detailed model specifications will be outlined in the statistical analysis plan

testing effects on the dependent variables; and helps to avoid the need for multiple comparisons adjustments such as the Bonferroni correction (Snijders and Bosker, 2012, p.283)**.** Significantly, the second and third advantage will potentially allow for stronger conclusions to be drawn, and additionally the third advantage will provide us with the opportunity to test the relationship between the score on the proposed financial knowledge and understanding instrument and the GCSE mathematics score.

Software

All models will be estimated via Bayesian inference. While Bayesian statistics is far too large a literature to describe here, it is based on personal rather than long-run interpretations of probability. Problems are divided into immediately available quantities and those needing to be described probabilistically. We describe our prior information via a full probability distribution and then sequentially update this knowledge with new data describing the resulting posterior quantities via quantiles or intervals rather than point estimates (see Browne, 2015; Gill, 2015). Using Bayesian inference provides several advantages for this evaluation – parameters are treated as random– an assumption far more suitable to social science applications where we would expect treatment effects not to be permanently fixed; more interpretable parameter estimates along with clear statements of uncertainty; more accurate estimates of variance parameters – particularly when the numbers of groups are small; missing data is handled seamlessly within the model; and finally results which set up further investigation via the sequential learning mechanism discussed above.

As we have little prior information from previous trials, we deliberately use diffuse/weakly informative priors to represent our ignorance on the size and variance of the treatment effects. Therefore, we deliberately use less prior information than we have available to allow the new data to dominate the overall results – this will be discussed further in the statistical analysis plan. We use the software STAN to fit the Bayesian models. Post-estimation processing and analysis will be carried out in R. As an additional check on the results we will also fit the analytical models using R2MLwiN/lme4 to provide consistency for the University of Durham's meta analyses and using the EEFanalytics package within R (making minor changes to priors used where appropriate). Where possible, missing data will be handled within the model where it is treated as another random parameter and estimated conditional on the observed data, but where there is significant missingness across multiple covariates we will use multilevel multiple imputation.

Process Evaluation

We will conduct a process evaluation that, in addition to focusing on the fidelity of implementation, aims to provide formative feedback in order to inform wider implementation and future scaling-up. The research questions for the process evaluation are detailed above (RQs 9 through 18).

Given the broad scope of the process evaluation, a variety of approaches will be needed targeted at specific aspects of the process, including existing data / records (where possible), questionnaires, observations and interviews, as follows:

Teacher records of how frequently real-world and personal finance contexts are used in maths lessons: To investigate fidelity, adaptation and quality of delivery**.**

Student perceptions of how, and how frequently, real-world and financial contexts are used in maths lessons in a sample of intervention and control classes: To investigate (i) fidelity, and (ii) enable statistical modelling of the relationship between dosage and outcomes. An earlier validation study indicates that student perceptions of the use of contexts and representations in class can be efficiently measured using a relatively short rating scale-based survey (Nitz et al, 2014). We will build on and (where possible) adapt this instrument to focus on the use of real-world and financial contexts. We will administer this using a paper-based machine-readable format in Summer 2018 to a subsample of at least two classes from each intervention (one taught by a core teacher, the other by another non-core Maths in Context teacher) and control school. This will enable us to additionally compare the use of real-world and financial contexts by core and non-core teachers taking part in the intervention.

Planned and actual attendance at the Maths in Context PD sessions: To investigate fidelity, we will collect developer records and registers of attendance by intervention school teachers for all external PD sessions.

Developer Professional Development Planning: The developer's planning and documentation for PD sessions (presentations, handouts and agendas) will be collected in order to provide information on the aims, approach and intended content of the PD sessions. This will be used to inform subsequent collection and analysis.

Observation at PD Sessions: We will observe two external training sessions, two consultant mentoring visits and at least one in-school cascade training sessions, in order to investigate (i)

how content about programme implementation is delivered and how teachers respond, (ii) how content about Maths in Context materials and desirable pedagogies is delivered and how teachers respond, and (iii) how information about programme management, Maths in Context materials and desirable pedagogies and information about mentoring strategies is delivered to mentors and how they respond. This will enable us to provide guidance and formative feedback on the wider implementation and scale up.

Interviews with the Maths in Context PD team: We will conduct short telephone interviews with at least three members of the development team and two Maths in Context accredited trainers responsible for delivering the PD sessions. This will provide evidence relating to the planned and actual content of the PD sessions and the intervention more generally.

Surveys of intervention teachers' financial capability: To understand the extent to which the intervention improves teachers' own knowledge and understanding of finance issues and financial capability as well as how this knowledge and understanding impacts on implementation and outcomes. We will survey all lead and other teachers involved in the intervention using adult version of the MAS financial knowledge and understanding instrument via the Bristol Online Survey (subject to validation).

Surveys of Intervention Participants and Schools: We will survey all intervention school teachers at the end of the PD in order to investigate their perceptions of the PD and their initial views on the YE Maths in Context programme. We will survey all participants and schools through short focused questionnaires that can produce quantifiable data in June 2018 (at the end of the first main year) and in June 2019 (at the end of the full intervention): a school questionnaire to be completed by head teachers or deputies relating to school leaderships' views on the efficacy and (expected) outcomes of the programme, a questionnaire relating to management and implementation to be completed by intervention teachers. Intervention teachers will be asked to complete a log of all instances of implementation that they consider they have applied in their teaching. These will provide data about fidelity and quality of the intervention in general.

We will survey pupils involved in all implementation cohorts to ascertain their perceptions of use of Maths in Context, materials and teachers' developing pedagogies.

We will ensure that these specially-developed instruments are easy and quick to complete and will either be delivered online administered via the Bristol Online Survey or will be paper-based but, as far as possible, using a machine readable format.

An additional survey to all intervention schools would allow for a more robust examination of this variability together with stronger evidence on the underlying factors. We will review whether this is desirable during 2017 and, if so, the protocol would be updated.

In-depth Case Studies of Intervention Schools: In order to explore implementation (and particularly constraints/affordances to implementation) we will conduct in-depth case studies in five intervention schools, purposively chosen to provide insight into diversity in practices. These case studies will be conducted over 5 days on two occasions (during autumn/spring 2017/18 and during spring/summer 2019). The later visit will enable us to collect participants' reflections on the intervention over the two years of Key Stage 4. We aim to organise initial case study research visits to coincide with mentor interventions in schools. During data collection in intervention schools, we will pay particular attention to issues relating to school and departmental management of the intervention, the use of the mentoring process and how this is managed.

Planned and actual implementation: In order to investigate fidelity to the intervention, we will collect evidence on the planned and actual implementation. Of particular importance in this regard are issues of a department's collaborative planning, scheme of work, sharing of resources, timetabling, assessment protocols etc. We will monitor teacher record-keeping of pupil progress and identify what if anything teachers do to focus specifically on the learning of financial capability and its interaction with mathematics.

Survey of usual practice: We will conduct a short survey of all schools' usual practice relating to the use of financial and related contexts in mathematics pre-randomisation and at the end of 2017. The survey will be targeted at Heads of Mathematics, will take no more than 10 minutes to complete and will be administered online via the Bristol Online Survey.

A detailed outline of the data collection for the process evaluation will be added as an Appendix to this protocol.

Cost Evaluation

We will follow the June 2016 EEF Guidance on Cost Evaluation in estimating the costs of the delivery of the intervention. We will collect cost data from the developer via a short interview and a pro-forma. In addition, we will collect data on costs incurred by schools through the process evaluation. Using evidence collected during the process evaluation, will estimate the costs of the cascade training, but will report this separately to the direct, marginal costs. As per the EEF guidance, we will report 'time' required separately to other costs.

Data management and analysis

Storage, backup and security of data

Our project data will be held locally and for this purpose we have the use of a secured partition on the University of Nottingham's research network drive with access provided to the four evaluation team members and our administrative support, as well as specific IT support personnel. Access to the network drives is tied to researchers' workstations which are situated in private locked offices. The terminals are all password-protected and authenticated through the University's centralised Information Access Management system. Remote access, where permissions allow for it will only work through remote desktop which has the same password and authentication requirements.

Primary pupil testing data and NPD secondary linked datasets will be held securely using AES256bit encryption on the research network drive and on the University of Nottingham's High Performance Computing cluster which can only be accessed from specific workstations in the School of Education. However, anonymised versions of qualitative data (interviews, observations etc.) may be stored on portable media (including laptops) and on cloud systems (e.g. Dropbox) without password protection to enable coding and analysis. Young Enterprise documentation will only be password protected where individuals and schools are identified.

Data protection, rights and access

The NPD data has strict rules in place with regards to privacy given the extensive personal details available with location, demographic and educational achievement data included. The descriptive analyses provide the greatest risk of undermining the privacy of participants and we therefore intend to avoid identifying individual schools where participants could be

identifiable. Inferential statistical analyses present far less risk as using Markov Chain Monte Carlo (MCMC) based samples mean that it is impossible to identify individuals.

Transparency, preservation, sharing and licensing

The long-term responsibility for archiving the data will rest with the Education Endowment Foundation and its partners. On the completion of the evaluation, the data and supporting documentation will be released to the Foundation, as well as the developers.

The research will make extensive use of open-source software – particularly R, but also Bayesian Inference software such as MLwiN and STAN. We therefore prefer to work with non-proprietary data formats such as tab-delimited data files, SQL databases, as well as ASCII style files for data analysis scripts and creating research outputs. However, some analyses will no doubt be tied to closed-source software. Where appropriate, we will create additional versions of the files in non-proprietary formats to promote long-term archiving.

Any academic publications which come from the work carried out during the evaluation will release supporting materials through Harvard's Dataverse repository. This is an open source data repository which provides the advantages of specifically organising data curation on the basis of published academic outputs. It allows the curation of replication code and documentation, provides controls over data access, and importantly provides versioning control for updates to material over time. All material which enables a successful replication (with the inclusion of synthetic versions of the data – to protect the identity of the participants, but maintain the statistical properties of the datasets) will be preserved in the repository.

Sensitive data such as from the linked National Pupil Database have a short period of access. Any sensitive data will be effectively destroyed. If the computer is being decommissioned at the same time as which the data is deleted, then the deletion will be carried out by the University's third party PC disposal partner, Stone. Otherwise it will be completed by IT Services staff. Data that is stored on hard copy (paper) will be shredded on-site prior to disposal. Retention of replication material will be permanent, although much of the original secondary data (linked NPD data) kept locally will need to be destroyed within a few months of the end of the project. Our analyses will use open source software such as R and so will need little preparation beyond appropriate descriptions within the script files. Data source documentation will need to be prepared, however, to enable researchers to be able to request data from EEF

and the planned deposit of material UK Data Archive to enable replication and follow on analyses.

Adherence

Adherence will be reviewed at the evaluation's progress management meetings and on request from DfE (for NPD related data) or the Education Endowment Foundation.

Ethical Issues

To ensure that the study has appropriate ethical oversight, each of the participating institutions – Young Enterprise and the University of Nottingham – will ensure that their respective ethics committees, or equivalent, have reviewed and granted approval for the intervention to go ahead. The processing and archiving of personal data for the project has been reviewed in the light of the General Data Protection Regulation (GDPR), which comes into force in May 2018. Data will be processed for public task purposes as per condition 6(1)e of the GDPR.[5] Because the basis for data processing has changed, parents / carers will be informed of this change in early June 2018 and will be given the option to object and withdraw their children's data from the project and any data processing. Teacher consent will be dealt with through opt-in consent for both the observations and interviews.

The trial is registered with ISRCTN (www.controlled-trials.com) reference number: ISRCTN58590757.

Personnel

University of Nottingham (UoN) / UCL Institute of Education (IOE)

The Evaluation team is made up of:

1. **Professor Geoff Wake (UoN)** will jointly direct the project and will lead on the development/pilot phase, the process evaluation and will contribute to the impact evaluation and validation of measures.

[5] The public task condition is similar to the public nature condition in Schedule 2 of the Data Protection Act 1998, the previous basis for the processing of data. For further information, see: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/public-task/

2. **Professor Jeremy Hodgen** (IOE) will jointly direct the project, will contribute to all aspects of the project and will lead on the impact evaluation and the validation of measures.
3. **Dr Michael Adkins** (UoN) will conduct the randomisation procedures, quantitative analysis and contribute to all aspects of the study, including drafting the evaluation protocol and report-writing.
4. **Professor Shaaron Ainsworth** (UoN) will advise on aspects of experimental design and will lead on the development of instrument measuring student perceptions of how, and how frequently, real-world and financial contexts are used in maths lessons.
5. **Sheila Evans** (UoN) will be responsible for the day-to-day management of the evaluation, maintaining contact with schools and undertaking the fieldwork and analysis relating to the process evaluation. She will assist with report-writing.
6. **Dr Dietmar Küchemann** (UoN) will contribute to the post-16 work.
7. **Kanchana Minton (UoN)** will be responsible for day-to-day administration, including assisting the team to maintain contact with schools.

Young Enterprise

The Young Enterprise team, acting as the developer, is made up of:

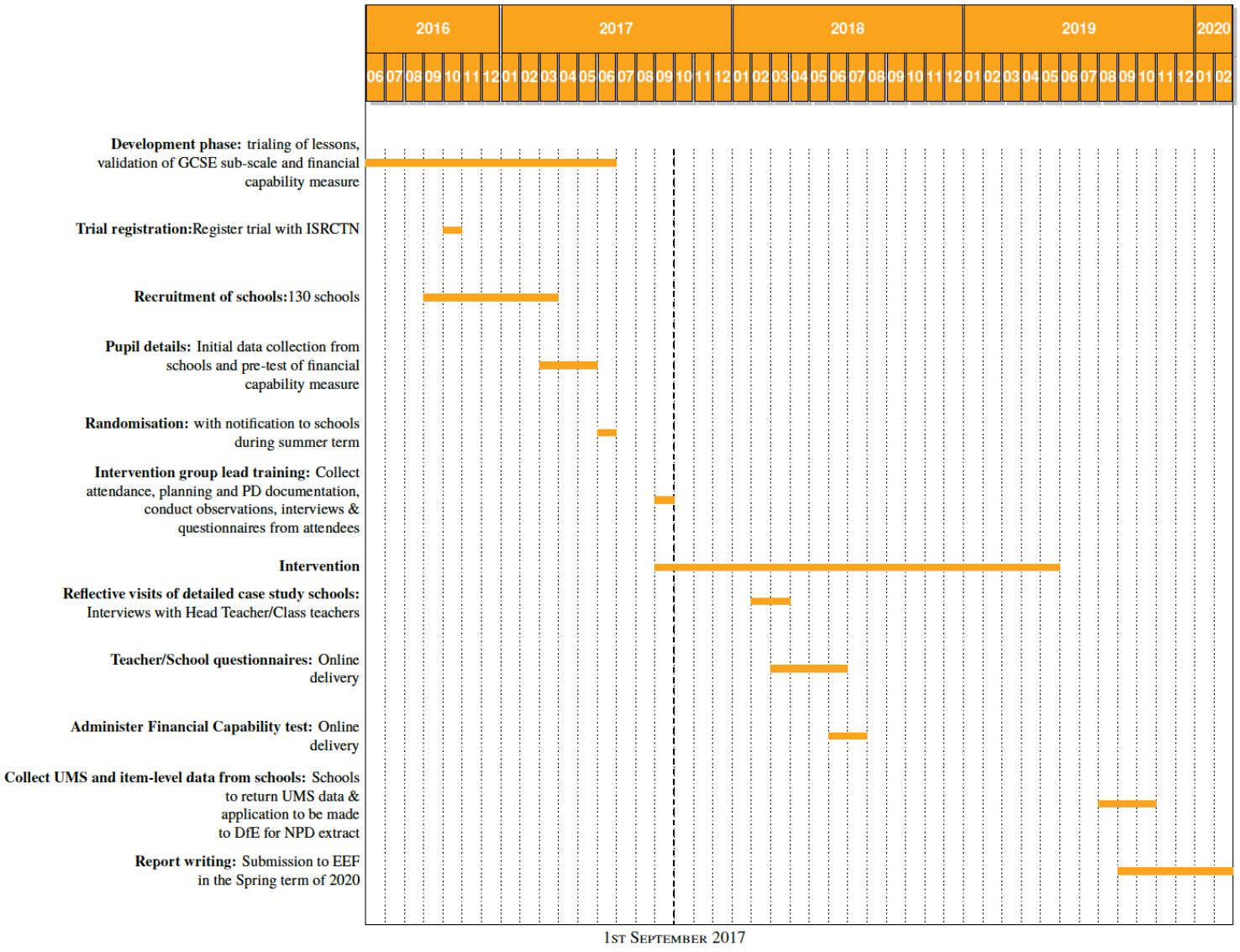
1. **Russell Winnard,** Head of Educator Facing Programme and Services
2. **Liz Booth,** Programme and Services Manager
3. **Ali Wakefield,** Project Manager
4. **Elle Craddock**, Project Coordinator
5. There will be an additional Project Coordinator joining the programme early in 2017.

Risk Analysis

We summarise our assessment of the main risks in the table below

Risk	*Likelihood*	*Impact*	*Action*
Failure to recruit schools	Low/Moderate	High	Involve developers and local / regional advocates in recruitment process. Run local / regional recruitment events together with a series of online webinars. Establish timeline for recruitment. Over recruit schools and set up a minimum recruitment target.
Attrition of schools and pupils	Moderate	Moderate	Over-recruit schools and pupils. Provide appropriate (and significant) financial incentives to control schools. Maintain contact with schools with through post-intervention period until GCSE data is returned.
Loss of staff	Low	Low	The School of Education at Nottingham has a large staff team so team members can be replaced.
Fidelity to intervention	Low	Low/Moderate	Provide feedback to development team during development / pilot phase. Monitor through process evaluation
Trustworthy and robust evaluation evidence	Low	Moderate/High	Robust process evaluation including triangulation of evidence and a focus on social validity in addition to fidelity
New GDPR may affect data processing and archiving	Low	High	Legal basis for processing of data reassessed April 2018. Advice provided to schools.

Evaluation Timeline



References

Atkinson, A. (2008) Evidence of impact: An overview of financial education evaluations, FSA Consumer Research Working Paper No. 68. [url] http://www.walcotfoundation.org.uk/uploads/1/7/2/2/17226772/fsa_overview_fin_education_2008.pdf

Bagwell, S., Hestbaek, C., Harries, E., Kail, A. (2014) Financial capability outcome frameworks, NPC [url] http://www.thinknpc.org/publications/financial-capability-outcome-frameworks/

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, *84*(2), 191-215.

Brown, M., Askew, M., Hodgen, J., Rhodes, V., Millett, A., Denvir, H., & William, D. (2008). Individual and cohort progression in learning numeracy ages 5-11: Results from the Leverhulme 5-year longitudinal study. In A. Dowker (Ed.), Mathematical Difficulties: Psychology and Intervention. Oxford: Elsevier.

De Jong, R., & Westerhof, K. J. (2001). The quality of student ratings of teacher behavior. Learning Environments Research, 4, 51e85.

Drake, P., Wake, G., Noyes, A. (2012) Assessing 'functionality' in school mathematics examinations: what does being human have to do with it? *Research in Mathematics Education* 14(3), 237-252

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. (2014) *Bayesian Data Analysis*, 3rd edition, Boca Raton, FL: CRC Press.

Gelman, A. & Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge: Cambridge University Press.

Hodgen, J. & Marks, R. (2013). *The Employment Equation: Why our young people need more maths for today's jobs.* London: The Sutton Trust

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

Huston, S.J. (2010). Measuring financial literacy. *The Journal of Consumer Affairs* 44 (2), 296-316.

Kaminski, J. A. Sloutsky, V. M., & Heckler, A.F. (2008). The advantage of abstract examples in learning math. *Science, 320*(5875), 454-455.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge, MA: Cambridge University Press.

Lesh, R., & Zawojewski, J. (2007). Problem solving and modeling. In F. K. J. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 763-804). Charlotte, NC: Information Age.

Money Advice Service (2013) *The Financial Capability of 15-17 year olds*, London: Money Advice Service

Money Advice Service (2015) *Financial Capability in the UK 2015: Initial results from the 2015 UK Financial Capability Survey*, London: Money Advice Service

Personal Finance Education Group. (2015). *Maths maters: usinf financial contexts in secondary mathematics*. London: PFEG.

Nitz, S., Ainsworth, S. E., Nerdel, C., & Prechtl, H. (2014). Do student perceptions of teaching predict the development of representational competence and biological knowledge? *Learning and Instruction, 31*, 13-22.

Noyes, A., Drake, P., Wake, G., & Murphy, R. (2010) Evaluating Mathematics Pathways: Final Report, [url]: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/182076/DFE-RR143.pdf

OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, OECD.

OECD (2014), *PISA 2012 Results: Students and Money: Financial Literacy Skills for the 21st Century* (Volume VI), PISA, OECD.

O'Prey. L., & Shephard D. (2014) Financial Education for Children and Youth: A Systematic Review and Meta - analysis. Aflatoun Working Paper 2014.1C. [url] http://www.aflatoun.org/evaluation

Pampaka, M., Kleanthous, I., Hutcheson, G., & Wake, G. "Measuring mathematics self-efficacy as a learning outcome." *Research in Mathematics Education* 13, no. 2(2011), 169-190.

Raudenbush, S.W. et al. (2011). Optimal Design Software for Multi-level and Longitudinal Research (Version 3.01) [Software]. Available from [url]: http://hlmsoft.net/od/

Vehtari, A., Gelman, A., Gabry, J. (2015) Efficient implementation of leave-one-out cross-validation and WAIC for evaluating fitted Bayesian models, [url] http://www.stat.columbia.edu/~gelman/research/unpublished/loo_stan.pdf

Wagner, W., Göllner, R., Helmke, A., Trautwein, U., & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: dimensionality and generalizability of domain-independent assessments. Learning and Instruction, 28, 1e11. [doi] http://dx.doi.org/10.1016/j.learninstruc.2013.03.003.

Wake, G. (2011). Introduction to the Special Issue: Deepening engagement in mathematics in pre-university education, *Research in Mathematics Education*. 13, 109-118.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary educational psychology*, *25*(1), 82-91.