Evaluation protocol
ScratchMaths (revised May2016)

Education Endowment Foundation

Amendments

1. Amended the evaluation summary table to reflect recruited numbers.
2. Made formatting changes, including updating the references to IoE to University College London IoE (from IoE) on first use and change in tense where activity has now occurred.
3. Page 3. Added a glossary at the start of the description of the intervention.
4. Page 3. Updated number of schools and classes from planned to actual numbers recruited and allocated.
5. Page 3. Updated the detail of follow up professional development: the number of days and twilights and also the number of clusters (hubs) and the teachers requested to attend PD (from one Y5 plus computing coordinator to two Y5 teachers with a computing coordinator as an alternate if necessary)
6. Page 5. Amended research question 6 from: 'What conclusions can be drawn about the relationship between mathematical thinking and computational thinking?' to 'what conclusions can be drawn about the relationship between mathematical thinking and computational thinking *from the quantitative analysis*?' This is to make clear the limits of what SHU will report on.
7. Page 6. Updated details of the number of design schools involved.
8. Page 7. Updated details of the CT measure development, including that correlation with KS1 will be established as well as KS2, and the number of pupils involved.
9. Page 7. A table has been added that gives the number of forms in recruited schools.
10. Page 7. Participants and recruitment updated with actual numbers.
11. Page 9. A new section added explaining the stratified randomisation process using propensity scores and a table showing the baseline school-level balance achieved following randomisation.
12. Page 9. Timing of the CT test updated and change to moderation process. Previously it had been planned that a small number of schools would be visited whilst the CT test was being administered. However, feedback on CT test administration will be obtained through surveying, the observations would not lead to generalisable conclusions about administration in other schools (due to observer effects) and resource will be better used to reduce attrition in the number of schools administering the test.
13. Page 9. Clarification that any *comparative* analysis of data of the CT test will be withheld.
14. Page 11. Analysis plan updated
15. Page 12-13. Description of fidelity measure added for Wave 1 Year 5 pupils additional detail added to the process evaluation as to how data related to the fidelity measure will be collected.
16. Page 15-16. Updates to information on IoE and SHU teams.
17. Page 17. The SHU Gantt chart updated to indicate revised timing of the surveys.

Evaluation Summary	
Age range	Primary (9 to 11 years old)
Number of pupils	4,400
Number of schools	110
Design	Clustered randomised controlled trial with randomisation at the school level
Primary Outcome	Mathematics KS2 attainment

Background

Mathematics and programming in schools have a longstanding and intertwined history. Programming in schools has been shown to have the potential to develop higher levels of mathematical thinking in relation to aspects of number linked to multiplicative reasoning; mathematical abstraction including algebraic thinking as well as problem solving abilities (Clements, 2000). More recently, attention has been paid to defining 'computational thinking' (CT) (McMaster, Rague & Anderson, 2010; Wing, 2008). Computational thinking could be considered to be a relative, or part of the 'family' of different aspects of mathematics thinking (Wing, 2008). This relationship helps to explain why programming and computer-based mathematical instruction have been found to have a positive effect on both student attitudes and on attainment in mathematics. Such effects are particularly apparent for students with attainment that is low relative to peers (Lee, 1990).

In the first phase of its introduction to the school curriculum, programming in schools was often developed by enthusiasts who would in many cases be located in mathematics departments, or, in Primary school contexts, would identify themselves as mathematics specialists. However, the introduction of Information and Communications Technology (ICT) as a National Curriculum subject has led to programming in schools being deprioritised both in relation to computing and in the mathematics curriculum. Recent policy and curriculum changes mean that there is a renewed focus on computing being (re-)introduced into Primary schools (DfE, 2013; Furber, 2012).

New programming languages, such as “Scratch”, and tools (for example, Raspberry Pi) have been developed since much of the research in computing and mathematical learning was undertaken. Scratch is a freely available programming and authoring tool developed for educational purposes at the Massachusetts Institute of Technology in the early 21st Century (Monroy-Hernandez & Resnick, 2008). Scratch is based on graphical programming blocks that can be assembled to create programs. The appearance of the blocks are related to their function and meaning. Scratch is highly interactive allowing for changes to be made to programmes whilst they are running. The scripting area is modelled on a physical desktop to encourage 'tinkering'. Scratch is designed to interface with multimedia allowing for applications that are meaningful to learners (Resnick et al. 2009).

Whilst Scratch is a development of earlier programming languages designed as learning environments, it represents a significant development. Thus, there is both an opportunity and a need to design and evaluate curricula and professional development programmes that can maximise the benefits of programming for students' mathematical thinking and attainment in the current context. This research assesses both quantitatively and qualitatively, how to design materials for students and teachers that directly address the learning of the skills needed in computational thinking and mathematical attainment.

Intervention

Glossary

Control schools: Wave 2 schools (defined below).

Design schools: schools involved in the design phase of the project.

Design year: school year 2014/15.

Intervention schools: Wave 1 schools (defined below).

Intervention year one: School year 2015/16, the year in which ScratchMaths was implemented with Y5 in Wave 1 schools.

Intervention year two: School year 2016/17, the year in which ScratchMaths was implemented with Y6 in Wave 1 schools.

Wave 1: Schools who attended ScratchMaths PD in summer 2015 and 2016 and implementing ScratchMaths with Y5 classes in 2015/16 and with Y6 in 2016/17.

Wave 2: School attending ScratchMaths PD in Summer/early autumn 2016, and implementing ScratchMaths with Y5 only in 2016/17

Description

ScratchMaths is an efficacy trial that will develop and evaluate the effect of learning to program on understanding computational thinking and mathematical thinking. It will compare the performance on selected tasks of students' mathematical thinking with those who do not engage with the materials and programming activities. One cohort of students in about 55 schools (110 classes) will engage in a two-year programme, the first year focusing on computational thinking and Scratch programming, and the second year on the learning of mathematical thinking through engagement with specially designed Scratch-mathematics curriculum and tasks. These are aligned with both the Primary Computing and Primary Mathematics National Curricula.

In each year of the intervention, two teachers in each school will be trained in the use of Scratch programming and the curriculum materials. Training takes place over two days. In intervention year one, the target group of teachers was 2 teachers in each school, if possible two Year 5 (Y5) teachers, who will be teaching the ScratchMaths curriculum the following year. Where they were unable to attend or it had not yet been confirmed who the Y5 teachers would be for the next school year, then one or more alternatives were asked to attend, such as the Computing Coordinator, who could share the training with other teachers in the school at a later date, alternatively another class teacher (Year 5 or Year 6). In year two, the target group where possible are two Year 6 (Y6) teachers, who will be trained to continue the approach with the same cohort of children.

In addition to the face-to-face teacher professional development (PD), an online community provides support and some optional light-touch collaborative support between nearby schools, with two additional optional half-day PD sessions per year.

ScratchMaths materials are organised into three **modules (per year)**.

Each module consists of a number of **investigations** designed to last in the range 50-70 minutes (which in reality is at least 2 lessons with technical setup etc.).

An investigation consists of core activities having certain steps which are designated as extensions as well as some full extension activities.

In addition there are one-off **challenges** that are available that are additional activities and entirely separate from the core materials.

Teacher module materials include notes on using the materials, vocabulary and concepts, links to primary national curriculum, class discussion points and example Scratch scripts.

In addition on the **ScratchMaths website** (http://www.scratchmaths.org/wp-login.php) teachers can also access PowerPoint presentations that can be used to support the use of the investigations, short videos related to some activities, and can download starter code for Scratch projects.

Schools were recruited in 7 geographical clusters. Compulsory professional development activities are led by the University College London Institute of Education (IoE) development team supported by local hub leads. The local hub leads will also lead three of the four optional half-day PD sessions over the two years (with the UCL IOE team leading on only the first of these).

Prior to the implementation of the intervention, the Institute of Education (IoE) undertook a design phase and SHU developed a measure of computational thinking based upon the IoE model of the construct by developing or adapting questions used in an international on-line project: Beaver computing (see below).

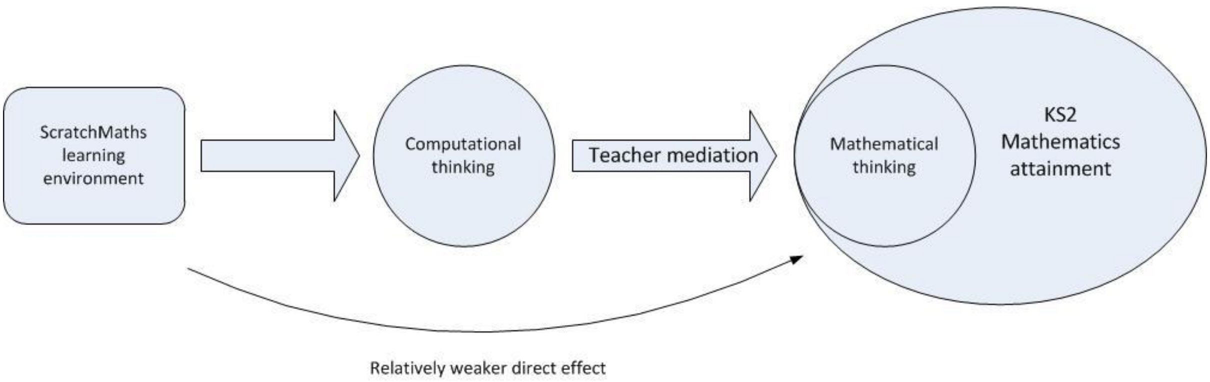
The intervention has a number of features of effective professional development: connecting work-based learning and external expertise; potentially rich professional learning opportunities; collaborative learning, the creation of professional learning communities between schools; and a clear focus (Stoll, Harris & Handscomb, 2012).

Theory of change

The diagrams, below, are an initial simplified theory of change of the proposed intervention effect.

Figure 1 represents a simplified model of the hypothesised relationship between ScratchMaths, computational thinking, mathematical thinking and KS2 mathematics attainment at the pupil level. Two aspects of simplification are the linear nature of the model without the representation of possible feedback loops and the separation of computational thinking and mathematical thinking as completely distinct. An aspect of the evaluation during the design phase will be to explore and describe in more detail the theory of change underlying the project, including feedback loops, in order to inform the process evaluation and potentially sub-analysis in the evaluation of impact through the RCT. This theory of change will be reported in the final report.

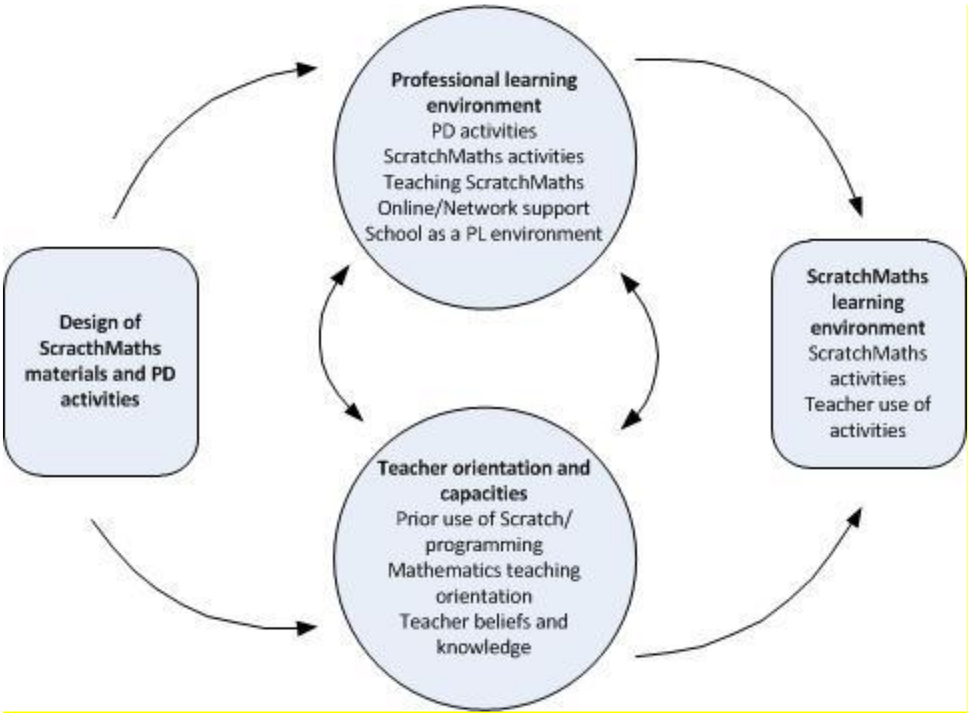
Figure 1 ScratchMaths and student learning



An important feature of the theory of change is the importance of teacher mediation. Whilst engagement in Scratch programming may lead to improvements in computational thinking, the direct effect on mathematical thinking is likely to be relatively weak. In order to enhance changes in mathematical thinking, teacher mediation is required. Teacher mediation refers to the role of teachers in making connections between computational thinking and Scratch programming and mathematics. This underlies the intervention design in which the main focus in Year 1 is on enhancing teaching in relation to programming and computational thinking, and developing materials and activities towards that goal. In Year 2 the main focus is linking programming learning and computational thinking to mathematical understanding. The project design is focused on developing mathematical thinking, which is one aspect of the capacities and knowledge that is assessed through KS2 Mathematics tests.

Figure 2 is an initial model of how the professional development and curriculum materials may lead to teacher change and create the ScratchMaths learning environment.

Figure 2 ScratchMaths as a professional learning system



However, the way in which professional development and teachers' engagement in curriculum innovation leads to changes in knowledge, practice and beliefs is more complex than this. One aspect of this complexity is due to the interconnection of the professional learning environment, the teacher orientation to intervention as well as their capacities. Secondly, as can be seen from the diagram, the particular ScratchMaths learning environment a teacher instigates is itself an aspect of the teacher's professional leaning environment. The model is of a set of nested systems (Opfer and Peddar, 2011). A more developed theory of change will be developed and explored during the project.

Research plan

Research questions

1. What has been the effect of the intervention on the development of pupils' mathematical skills as measured by a randomised control trial?
2. How can computational thinking be measured?
3. What correlation exists between measured computational thinking and mathematics attainment?
4. What has been the impact of the intervention on the development of pupils' computational thinking?
5. What conclusions can be drawn about the relationship between mathematical thinking and computational thinking from the quantitative analysis?
6. To what extent does the design and delivery of curriculum materials and PD and the associated materials fit with the current knowledge base on effective PD in relation to mathematics teaching/computing?
7. What are the teachers' views on the effectiveness of the PD?

8. Were there any barriers to implementing Scratch, or where there particular conditions that needed to be in place for it to succeed?
9. In what ways can the PD delivery and materials be improved?
10. What issues of fidelity occur during the trial, and what is the security of the trial outcome (taking into account any use of Scratch in control schools?)

In addition to these research questions there will be an exploration through the process evaluation of the scalability of the trial.

INTERVENTION PILOT AND DEVELOPMENT

During the first year of the project (2014/15), IoE undertook a developmental year in order to design and trial materials for the intervention (both Y5 and Y6 materials), and associated PD. It was planned that five 'design schools' would be involved. In the event four were fully involved and a further school had partial involvement. The design schools represented a range of school types and previous levels of engagement with Scratch programming. The pilot phase involved:

- Review of the literature & available materials
- Collation, design & trial with the design schools a package of materials (teacher, student and PD) in preparation for the main trial focused on computational thinking (Y5) and mathematics and computational thinking (Y6)

Sheffield Hallam University (SHU) was not actively involved in the trialling/design work, but was kept abreast of developments, and materials were shared to support the final design of the trial and process evaluation. The curriculum structure for both years of the intervention was developed prior to the commencement of the trial. Additionally all of the Y5 content was designed prior to the Wave 1 Y5 PD, with a finalised version of all materials available online prior to the Wave 1 Y5 teachers commencing the delivery of the Y5 intervention in September 2015. Similarly using the same initial structure, all of the Y6 content was designed prior to the Wave 1 Y6 PD, again with a finalised version of all materials available online prior to the Wave 1 Y6 teachers commencing the delivery of the Y6 intervention in September 2016.

DEVELOPMENT OF THE COMPUTATIONAL THINKING MEASURE

Parallel with development of the intervention during 2014/15, IoE supported SHU on the development of a computational thinking measure. A starting point for the development of the scale was to draw on tasks developed for an international computing challenge – the Beaver contest[1], an international initiative with the goal to promote informatics (Computer Science, computing) and computational thinking among pupils of primary, lower secondary and upper secondary stages. Each year the contest comprises of a set of tasks (items) of which some are directly related to computational concepts and computational practices.

Members of the SHU and IoE teams met in November 2014 to:

- develop a shared understanding of IoE's operational definition of computational thinking as used in the intervention design,
- and discuss examples of Beaver questions.

Following this, SHU then designed, developed and tested the CT measure independently. The CT measure design was undertaken in 4 steps:

1. Step 1. Initial design review of Beaver questions and other items by SHU leading to a first draft of the scale.

[1] Established in 2004, see http://www.bebras.org or http://www.beaver-comp.org.uk or http://www.ibobor.sk/ it is now run in 30 countries. In 2013, more than 720,000 pupils took part in the contest.

2. Step 2. Trial of test items in two schools (planned 60 Y5, 60 Y6 pupils; actual – 49 year 5 pupils and 66 year 6 pupils, 115 in total) statistical analysis of items/scale leading to a revised scale and test protocols.

3. Step 3 - test of revised scale plus focus on trial of protocols - 1 school. The plan was initially for 30 Y5 and 30 Y6, but it was trialled with 57 Y5, the focus on Y5 only considered more appropriate as a test of use in the main trial.

4. Step 4 - Full test to establish the correlation between CT and both KS1 and KS2 with 6 schools (planned 360 Y6 pupils, realised 231 pupils for which Unique Pupil Numbers and opt out consent obtained). The initial intended number was based on a model of a moderate correlation (0.3) and a criterion for significance of .05 and a power of 0.80 for the analysis for the FSM pupils (which would be expected to constitute approximately 1/5 of the total pupils in the sample (approximately 65 FSM pupils). The power of the analysis for the general population would naturally be higher. Further analysis will be undertaken using data retrieved from the NPD about the suitability of the achieved sample size.

Each school in the CT design process was offered the opportunity for two teachers to go to a one day and a separate half day professional development session run by Sheffield Hallam University on teaching computing in primary schools. There were 9 schools in total involved. The schools were recruited in the Sheffield City region and were not involved in either the IoE design process or the RCT.

Randomised Control trial design

The research design is a Clustered Randomised Control Trial (CRCT) with 3 levels: pupils are clustered into classes which are clustered into schools. Schools were randomised to either the treatment or a wait-list control group (see Randomisation section below for detail).

The trial used a wait list design with the Wave 2 (control) schools' Y5 teachers engaging in PD in the second year of the intervention (2016/17) and will be free to use materials with their Y5 pupils. However children in Y6 in the Wave 2 (control) schools will not be taught or access ScratchMaths materials in the second year in order that the control group is retained.

Participants and recruitment

The trial planned to involve the recruitment of approximately 115 schools. It was anticipated there would be drop out at the point of agreeing to trial protocols and the design aimed for 100 participating schools at the point of randomisation (50 intervention, 50 control). In the event, 111 were recruited and 110 allocated to the intervention and control conditions (see below). The trial was to involve, ideally two classes of Y5 pupils per school, and the same pupils when they progress into Y6. The target recruitment was for two-form entry schools wherever possible. Having the same number of Y5 classes would simplify UPN data collection and opt out consent as the whole year group will be involved. The aim for two form entries was relaxed in order to support recruitment and the distribution of classes in schools is given below

Table 1 Sample by number of forms

Number of forms	Intervention	Control	**Total**
1	5	2	**7**
1.5	1	0	**1**
2	43	43	**86**
3	6	8	**14**
4	0	2	**2**
Total	**55**	**55**	**110**

Table 1 data source: table compiled from number of forms indicated by information supplied by school to IoE, with missing data completed from Raise online as source for 10 of the 110 schools.

There is the possibility of a small dilution effect for the six 3-form entry intervention schools (though arguably a balancing concentration effect in 1 form entry schools). School size/number of forms of entry will be included in the intention to treat analysis model, which may give evidence of dilution effect. The issue can be further considered if there is an on-treatment analysis by excluding schools that are not 2 form entry from the follow up analysis.

Recruitment began during the design phase with the aim of identifying all schools by March 15th 2015. This was to allow for signing of Memorandum of Understandings (MOUs), collection of pupils Unique Pupil Numbers (UPNs) and National Pupil Database (NPD) data retrieval prior to randomisation. IoE aimed to identify 5 hub locations through discussion with National Association of Advisors for Computers in Education (UK) (NAACE). In the event, 7 hubs were identified that were geographically spread and had different profiles in terms of degrees of urban or rural contexts. Recruitment was undertaken by IoE with support from NAACE. Once schools had been recruited, SHU collected UPNs for the focus cohort (Y5 in 2015/16) and other school and teacher level data as needed.

SHU and IoE co-produced information including initial recruitment information, consent forms and MOUs for use with the schools. IoE supplies SHU with information on the recruitment process using the EEF participant flow diagram[2].

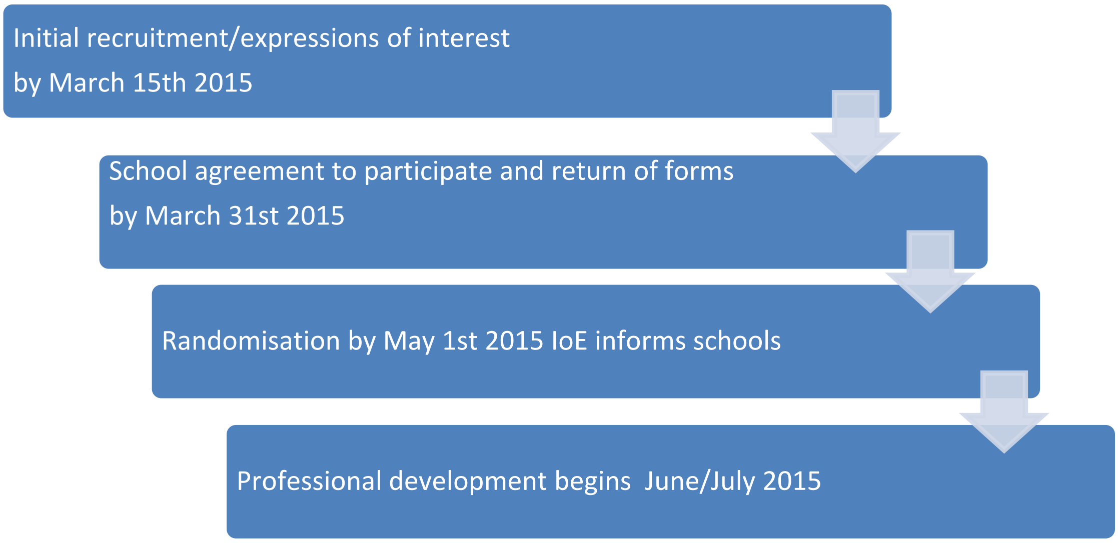
Schools were required to provide the following as a condition of being entered into the randomisation:

- MoU signed by the head teacher. The MoU included details of the requirements for the CT test in Summer 2016 and both IoE and SHU evaluation activities, as well as information the school will be expected to supply
- Information on who would be attending the PD events if allocated to the intervention group
- Summary information on any previous use of Scratch programming, or engagement with Beaver tests (June 2015)
- Pupil lists for Y5 including UPNs
- Confirmation the school has sent out the parent opt-out consent form.

Figure 3 describes the process of recruitment to the start of professional development with dates.

[2] Page 8 of https://educationendowmentfoundation.org.uk/public/files/Evaluation/2015_trial_reporting_template.pdf

Figure 3 Recruitment and allocation timeline



Randomisation

Using details from the 2013/14 Annual School Census (ASC) for all schools recruited to the Scratch evaluation, a stratification scheme was developed using propensity scores (Holmes, 2014). Propensity scores provide a practical way of drawing on a greater number of explanatory variables within a stratification scheme. A logistic regression model was used to generate (predicted probability or propensity) scores based on a KS2 attainment outcome variable[3] and seven explanatory variables[4]. Within each of the hub areas, the propensity scores were used to group schools into their 'nearest statistical neighbour' pairs. One school from each pair was then randomly selected into the intervention (or wave 1) group, the remaining school was allocated to the control (or wave 2) group.

The propensity score paired-school stratification approach required an even number of schools in all of the hub areas. This was not the case for three; an odd number of schools were recruited in both of the London and in the Somerset hub areas. The two (north and south) London hubs were combined into a single one with an even number of schools prior to progressing with pairing them up and randomisation. Within the Somerset area, the propensity scores identified one school to be very distinct from the remaining 16 schools. This school was then dropped and the remaining schools progressed to being paired and randomised to the intervention or control group.

In all, 55 schools were randomly selected to receive the Scratch programming intervention within the trial period (wave 1) and their 55 pairs were allocated into the control group, being offered the Scratch programming intervention after the trial has ended (wave 2). The school that was dropped prior to the randomisation stage will also be offered the intervention after the trial has ended, but will not be included within the impact analyses.

The reason for adopting propensity scores to provide a multivariate stratification scheme for this trial was to try to best ensure a good school-level baseline balance. A simple randomisation approach or a simpler approach to stratification would have been less likely to ensure this. As Table 1 shows, the school-level baseline balance achieved by this approach was good across the seven variables. Further analyses showed that there was a good balance between the samples within each of the hub areas.

[3] A binary outcome that identified whether the proportion of pupils within a school attaining a level 5 or higher in KS2 mathematics was greater than the median population value of 42% (=1) or not (=0).
[4] Explanatory variables - KS1 attainment, KS1 to KS2 progress in mathematics, school size, gender balance, %FSM, %EAL, %SEN.

Table 1: comparing the wave 1 and 2 school samples

	Wave 1 Sample (Intervention) n=55	Wave 2 Sample (Control) n=55
KS1 Attainment (Points Score)	14.9 (1.58)	14.9 (1.32)
KS1 to KS2 VA - overall	100.3 (1.15)	100.2 (1.18)
KS1 to KS2 VA - maths	100.4 (1.44)	100.3 (1.50)
% with level 5+ in KS2 Maths	42.3 (17.18)	41.7 (16.17)
%FSM (ever in last 6 years)	32.5 (20.61)	31.8 (17.49)
%EAL	24.1 (32.84)	30.2 (33.72)
% SEN of SAP	11.5 (7.24)	11.3 (8.06)

Balance at individual level and exact number of pupils will be determined once National Pupil Database (NPD) data is retrieved.

Outcome Measures

KS1 mathematics assessment data will be used as the pre-test, and will be retrieved from the NPD.

The primary outcome measure will be KS2 mathematics test scores at the end of Y6.

An intermediary secondary measure at the end of year 1 of the intervention will be a measure of aspects of computational thinking appropriate to the age range. This will be undertaken as an on-line test in May/June 2016 ideally prior to the 2016 PD sessions for both Wave 1 (intervention) and Wave 2 (control) schools. Measuring computational thinking at the end of Year 1 (2015/16) will address the intended outcomes of the first year of the trial where the focus is on computing and computational thinking. In addition, it will allow analysis at the end of Year 2 (2016/17), of the relationship (if any) between measured differences in computational thinking and the impact on mathematics attainment.

Testing of pupils will be staggered, this is to accommodate schools' access to IT facilities and the potential need for support with log-in for pupils and IT support by SHU. It is envisaged that the CT test will take place over a 2 week period. The date and time of the test will be recorded and any evidence of impact of staggering the tests will be analysed. All Y5s in the schools will be tested, scoring will be undertaken through coding, and therefore the assessment will be blind. Teachers will be responsible for invigilating the tests.

Outcomes of the comparative analysis of the CT measure will be withheld from all schools and IoE until summer 2017 when the trial is complete.

The theory of change (see Fig 1) identifies that the intervention is intended to impact on mathematical thinking (a component of mathematical attainment), as measured by KS2 tests. Further, there will be wide variation in how closely related aspects of ScratchMaths are to different aspects of the primary mathematics curriculum. A more fine grained analysis may be possible if a subset of KS2 tests questions were used. The KS2 tests framework[5] (Standards and Testing Agency 2014) would allow for this if data was available from the NPD. This would give a more sensitive measure of the impact of the intervention on specific areas of mathematics attainment. This could form part of an 'on treatment' analysis depending on the outcome of the trial and the availability of item by item analysis in 2017.

Sample size calculations

[5] Standards and Testing Agency (2014) Key stage 2 mathematics test framework: national curriculum tests from 2016. UTL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/295174/2016_Key_stage_2_Mathematics_test_framework.pdf

A power analysis was undertaken using the Optimal Design Software and the findings are presented in Table 1. The research design is a clustered randomised controlled trial with three levels (pupils are clustered into classes that are clustered into schools) with randomisation at the school level.

The power analysis presents the estimated Minimum Detectable Effect Sizes (MDES) for the primary outcome (KS2 maths attainment at the end of Y6 in summer 2017) and the secondary outcome (computational thinking test at the end of Y5 in summer 2016). For these MDES estimates, a statistical power of 0.8 is adopted, 110 schools were recruited and two classes per school will participate in the trial, with the average class size of 20 pupils.

A statistical power of 0.8 or 80% means that the research has been designed so that there is an 80% chance of detecting an effect (or difference) between the intervention and control group samples if a true effect exists. The MDES is the smallest effect size that the research design would be able to detect as being statistically significant with a statistical power of 0.8. For example, an MDES of 0.23 indicates that a difference in the outcome scores of the intervention and control group of 0.23 standard deviations (or greater) would be identified as statistically significant with a statistical power of 0.8.

It is estimated that 13% of the variation in KS2 maths attainment will lie at the school level, and 7% will lie at the class level[6].

Table 2 sets out the estimated MDES for the primary and secondary outcomes of the Scratch programming clustered RCT. MDES estimates for an outcome only analysis are shown. Additionally, MDES estimates for analyses that include KS1 maths as a covariate are also shown. For the primary outcome, the correlation between KS1 and KS2 maths is estimated at (r=) 0.77 based upon guidance from the EEF. For the secondary outcome, a more conservative correlation between KS1 maths and computational thinking is adopted (r=0.5)

Table 2: Estimated Minimum Detectable Effect Size (MDES) for planned analyses for Scratch programming clustered RCT

- **Number of schools: 110**
- **Number of classes (2 per school): 220**
- **Number of pupils (20 per class): 4,400**

Primary Outcome - KS2 Maths in summer 2017	
Outcome Only	0.23 standard deviations
Including KS1 maths covariate (r=0.77)	0.18
Secondary Outcome - computational thinking in summer 2016	
Outcome Only	0.23 standard deviations
Including KS1 maths covariate (r=0.50)	0.21

ANALYSIS PLAN

Descriptive analyses of measures at baseline and the two outcome stages will be undertaken. The baseline analysis will examine how similar (or balanced) the intervention and control group samples are in terms of factors such as KS1 maths attainment, gender and FSM. To assess how attrition impacts on this balance, the baseline analyses will be replicated on the sample with complete baseline and outcome measures. For example, the average KS1 maths attainment for the control and intervention group samples might be similar at baseline but significantly different amongst the subsample completing the trial[7].

Following descriptive analyses, a multilevel analysis will be undertaken for both primary and secondary outcomes to acknowledge clustering with schools and classes. For the main impact analyses, two analyses are

[6] The school level ICC of 0.13 is taken from the EEF guidance from analyses of NPD 2013-2014 and the class level ICC of 0.07 is estimated as being half of what is found at the school level (due to the wide spread practice of setting within primary mathematics).

[7] That is excluding pupils who might 'opt out' of participating in the trial.

planned. First, an outcome only analysis that will just include the dummy variable that identifies whether a pupil is in intervention or control group will be included - these analyses have higher MDES estimates but will be immune to any potential issues of bias resulting from missing KS1 maths data. Second, the analyses will include the KS1 maths covariate. It is these analyses that will be used to assess 'impact' of the Scratch Programming intervention on the primary and secondary outcomes and they will be undertaken as an 'intention to treat' analyses.

Following the main impact analyses, further exploratory analyses are planned. These exploratory additional analyses aim to explore whether the impact of the Scratch Programming intervention was stronger for particular pupil subgroups - specifically in terms of pupils classed as FSM and pupil gender. Multilevel analyses (outcome only and including the KS1 maths covariate) will be run separately for pupils classified as FSM and pupils not classified as FSM. Similarly, multilevel analyses will be run separately for males and females.

Finally, a fidelity analysis will be undertaken amongst the intervention school sample. Measures of fidelity were identified as part of the process evaluation and informed by IoE's design evaluation (see below).

For Wave 1 Year 5 pupils, elements of a minimum measure of fidelity are:

- Taught by a teacher from the Wave 1 school who attends at least two days equivalent of PD, where this is defined as any combination of Summer 2015 PD days or half day optional PD or substantial in school PD. An example of the latter would be through cascade by a teacher who attended any of the Summer PD missed and by extensive co-planning with a teacher who attended PD
    - Taught by teachers able to access appropriate resources including computers running Scratch 2.0 offline or with adequate internet access and typically a minimum 2:1 pupil to computer ratio for at least 1 hour per fortnight
- Pupils are taught at least some 'the core' activities from at least three investigations from each module
- Time spent on investigations taught is the minimum specified
- The teaching of ScratchMaths respects the progression built into the modules; so the order of modules and the order of activities are followed in general.

The minimum fidelity for the intervention in Year 6 will be similar. A combined fidelity measure for the intervention as a whole will be fidelity in Y5 and Y6.

Data on fidelity will be collected at the teacher level. Multilevel models will examine whether fidelity significantly accounted for variation in the outcome variables. Depending on the outcome of the fidelity analyses, additional 'on-treatment' analyses may be undertaken. For example, if engagement with the intervention (as captured by the measures of fidelity) was found to have a statistically significant impact on the outcome variable, an 'on-treatment' analysis would be included in addition to the main 'intention to treat' impact analyses.

PROCESS EVALUATION METHODS

IoE will conduct a 'curriculum design'[8] evaluation, and SHU will conduct a process evaluation. Regular contact between SHU and IoE will ensure that evaluation activities will be co-ordinated. The IoE's design evaluation will focus on how the curriculum materials are used by teachers and engaged with by students, as well as observed qualitative outcomes of this. In addition, teachers engagement' with professional development, and the outcomes of this, will be identified by IoE. The focus of the curriculum design evaluation will be to inform the refinement of the intervention and to understand its effects.

[8] The curriculum design evaluation refers to the project as a whole and not only to the design year.

The SHU process evaluation will examine how the intervention is conducted and also develop an independent view of the materials and the PD, including teachers' views on it. The process evaluation will validate the IoE's analysis of professional learning. It will also address issues of fidelity, validity of the trial and scalability of the intervention.

The process evaluation aimed during the design year: to provide an independent view on the process of the design of the curriculum materials and associated PD activities and the IoE's evaluation of these, and to provide guidance to the project team on ensuring that the intervention approaches, materials and training will be replicable and testable through a randomised control trial.

The process evaluation aims during the Intervention years are to evaluate the reliability and validity of any identified impact through a process evaluation that will identify issues of fidelity and scalability, in particular the barriers and necessary conditions for successful implementation, and to address the evaluation research questions. IoE will keep records of attendance at PD events and undertake initial data collection on technological prerequisites and the organisation of computing in schools. SHU will directly ask schools to identify class teachers of Y5 and Y6, so as to link with the attendance data. Y5 participating teachers will be surveyed in 2016, and Y6 participating teachers in 2017 to collect data on other aspects of implementation. Response items on the survey will be informed by IoE analysis of open questions on initial surveys and analysis of teacher interviews.

At a meeting in November 2014, SHU and IoE teams met so that SHU understood IoE's plan for the 'curriculum design' evaluation to inform the process evaluation and ensure there is no replication so that schools are not overburdened. SHU and IoE will continue to work to ensure that data collection is complementary. In addition, the teams will review and further develop the theory of change to inform the process evaluation.

SHU will draw on IoE's data and evaluation where possible, and provide an external view on the conclusions drawn from it, using this to inform an assessment of the validity and reliability of the RCT outcome.
The following data/materials will be required from IoE to support the process evaluation:

- The summary design evaluation report at the end of the design year
- Copies of training and curriculum materials for Y5 and Y6 teachers by Sep 2015 ,the start of 2015/16 (and any adjustments to the Y6 materials by the start of 2016/17)
- Information on recruited schools and teachers prior to randomisation: address, head teacher, chair of governors.
- Information before start of intervention, where possible, of the 2 teachers participating in the first year of the trial and the 2 teachers participating in Year 2. In addition base line information about existing use of Scratch should be collected.
- Attendance records of teachers at PD events
- Notes on notable issues related to delivery of PD
- Records of any issues that might affect fidelity: e.g. any changes to the teachers who deliver the interventions; record of teacher use of the materials (quantitative summary, this would include information on whether the materials are taught by someone other than the class teacher)
- Summaries of participation in on-line activity (quantitative data on who, when, frequency)
- Information on project costs.

The table below gives detail of process evaluation methods.

Table 3 Process evaluation

Year	Activity	Details
Pilot and design 2014/15	Review of IoE design evaluation	Summary outcomes of data collected during the design phase by IoE will be reviewed. This will inform design of the process evaluation tools. Collection of data on school and teacher profiles during recruitment phase.
Intervention with Y5 2015/16	Visit to two Professional Development events - first and second day of training in two separate hubs.	Observation of PD, informal discussion if possible with teachers and PD leaders - key foci fidelity in use of PD materials, the nature of the PD used.
	Telephone interviews with 10 teachers in intervention schools.	Semi-structured interviews focused on: experience of PD, key professional learning outcomes, use of curriculum materials, changes in practice. Key foci on fidelity in use of curriculum materials.
	Survey of all teachers in the intervention and control schools.	Collect fidelity data on implementation in Wave 1 schools. Collect data on any other practices/activities that might influence computational thinking and/or mathematics e.g. other interventions. For intervention teachers, evaluation of PD and curriculum materials, affordances and barriers to engagement. Identify any issues that might affect balance. Also use of online or additional support.
	Review of IoE design evaluation data.	Additional data sources on fidelity will inform design of evaluation tools for the following year.
Intervention with Y6 2016/17	Visits to 3 PD events (hubs not visited the previous year).	Observation of PD, informal discussion if possible with teachers and PD leaders - key foci fidelity in use of PD materials and the nature of the PD used.
	Telephone interviews with 10 teachers in intervention schools.	Semi-structured interviews focused on: experience of PD, key professional learning outcomes, use of curriculum materials, changes in practice. Key foci on fidelity in use of curriculum materials. Identify issues of school level professional learning community.
	Survey of all teachers in the intervention and control schools.	Collect fidelity data on implementation in Wave 1 schools. Collect data on any other practices/activities that might influence computational thinking and/or mathematics e.g. other interventions. For intervention teachers: evaluation of PD and curriculum materials, affordances and barriers to engagement. Identify any issues that might affect balance. Also use of online or additional support.

Personnel

Institute of Education

Professor Richard Noss: Principal Investigator

Overall Project Investigator and responsibility for delivery of all outcomes. Overseeing project to ensure recruitment, quality of provision and implementation, CPD. consistency of aims and methods for CT and for Mathematics, ensuring the design research and work in schools provides evidence of the key pedagogical and teacher factors that underlie any success.

Professor Celia Hoyles: Co-Principal Investigator

Particular responsibility for mathematics CPD; qualitative outcomes & reports, managing (with NAACE) the (complex) logistics of training in hubs, outward-facing liaison with schools, teachers, dissemination in general (e.g. with policy-makers).

Professor Ivan Kalas: Co-Investigator

Responsible for literature review (in particular Beaver texts published in Slovak or Czech languages), responsible for design & drafting all Scratch tasks, build a selection of Beaver tasks (used in previous years) and extend this selection by several Beaver-like tasks, validation of Beaver tasks, co-ordination of design evaluation.

Professor Dave Pratt: Co-Investigator

Responsible for literature review of a research with Scratch, responsible for final design & validation of Scratch tasks (& design of maths Scratch tasks, both with IK); aligning the design evaluation and the process evaluation, in relation to fidelity measures and school level data collection drafting qualitative reports of first intervention

Research Officer Laura Benton

Provisional: Responsible for day-to-day liaison with schools, data collection, analysis under supervision, implementation of training as designed for treatment & control, alignment of tasks and activities with school curricula and practice, assistance to teachers in CPD sessions for technical and organisational issues

Project Administrator Kim Parsons

Responsible for maintenance of databases, design of website and maintenance, organisation of school visits, liaison with external bodies, coordination of staffing.

IOE PhD students:

Piers Saunders and part time research officer

Johanna Carvajal

Naace

Mark Chambers: CEO

Responsible for recruitment & leading the design of training & the training team (for all treatment teachers and for control teachers)

Advisory Group members

Professor Janet Ainley, University of Leicester

Miles Berry, University of Roehampton

Joe Halloran, Lambeth City Learning Centre

Gillian Ingram, Manager, Camden City Learning Centre

Debbie Morgan, Director for Primary, National Centre for Excellence in the Teaching of Mathematics

Dr. Mary Webb, King's College London

Sheffield Hallam University

Dr Mark Boylan: Evaluation Project director/PI

Sean Demack: Lead statistician

John Reidy: CT measure design

Anna Stevens: Data management, analysis of pupil progress, trial management of the CT measure

Claire Wolstenholme: Project manager

Martin Culliney: Research Fellow

Ian Guest: Research associate, process evaluator

Professor Hilary Povey: Programming in Mathematics advisor and process evaluator

Phil Spencer: Computer Science in Primary advisor

Ian Chesters, Administrator will provide administrative support

Risks

Main issues or risks to the evaluation and how they would be addressed:

Risk	Assessment	Countermeasures and contingency plan
Intervention insufficient in scale and time to impact on pupil attainment	***Likelihood:*** medium ***Impact:*** High	Impact evaluation designed across 2 years to minimise this impact. Ensure that scale of intervention allows for an RCT of sufficient sensitivity and power
Developer fails to recruit sufficient schools for control and intervention	***Likelihood:*** medium ***Impact:*** High	Begin recruitment process at start of design year. Design year is a full year. If numbers below target, re-design RCT to have unequal number of intervention and control schools
Insufficient Beaver items or unsuitable items to construct CT measure	***Likelihood Medium*** ***Impact Medium***	Use items from other related psychometric scales
Delays in access to pupil data	***Likelihood:*** Low ***Impact:*** High	Sufficient lead in ensures sufficient time for access of base line data.
Curriculum and PD design evaluation during pilot year indicates that the intervention is not workable	***Likelihood:*** *low* ***Impact:*** High	Early identification of issues and dialogue with EEF to review continuation to intervention
Data protection and ethics issues	***Likelihood:*** Low. ***Impact:*** High.	Robust data protection and ethical procedures are in place at SHU and CEE and data sharing protocols will be established
Staffing issues: staff leaving/unavailable	***Likelihood***: Low (turnover low) ***Impact:*** High.	Succession planning has been built into team roles and additional capacity within CEIR to replace all roles.
Project management and partnership working issues	***Likelihood:*** Low ***Impact:*** High	Directors and project manager highly experienced in successful project management. Well-developed project procedures.

The trial will be subject to ethical approval by both SHU and IoE's ethics processes. NPD data will be subject to NPD protocols about data sharing.

Procedures are in place to comply with the 1998 Data Protection Act. No information about any identified individual will be reported or made available to anyone beyond the project teams. All data will be stored anonymously and securely. Consent forms, participant information, and digital recordings will be stored separately from transcripts and case reports. In disseminating findings, names of respondents will appear as pseudonyms, and any other potentially identifying data will be anonymised. Personal data is only stored on encrypted portable media in password-protected files (and only when absolutely necessary). To the best of our knowledge, both Universities confirm to the principles of ISO/IEC 27001

Timeline

SHEFFIELD HALLAM UNIVERSITY: EXTERNAL EVALUATION

The evaluation timeline below will be amended when more detailed information is provided on the project timeline and evaluation activities being undertaken by IoE

	May -Jul 2014	Aug-Oct 2014	Nov-Jan 2014-15	Feb-Apr 2015	May-Jul 2015	Aug-Oct 2015	Nov-Jan 2015-16	Feb-Apr 2016	May-Jul 2016	Aug-Oct 2016	Nov-Jan 2016-17	Feb-Apr 2017	May-Jul 2017	Aug-Oct 2017	Nov-Jan 2017-18
1. Evaluation of design and pilot															
Inception															
Collect and analyse internal evaluation data															
Process evaluation activity															
School MoU and data collection															
Development of CT measure: Design Wave 1															
Development of CT measure wave 2															
Development of CT measure wave 3															
CT correlation with KS2															
2. Main trial year 1															
Recruitment			Jan	Mar											
Randomisation															
2X visits to PD events					J/J										
X10 Telephone interviews with teachers in project															
Survey(s) of participants (control and intervention)															
CT Testing															
Retrieval and analysis of pupil data															
Review and analysis of internal evaluation data															
3. Main trial year 2															
3 X visits to PD events															
10X Telephone interviews with teachers in project															
Survey(s) of participants (control and intervention)															
Review and analysis of internal evaluation data															
4 Analysis and reporting															
Retrieval of pupil data															
RCT analysis															
Reporting															

IOE TIMELINE

Sept 2014- Apr 2015	Phase 1: Setup *Design 1*	• Project team develop the CT & ScratchMaths interventions and PD in conjunction with 5-London based 'design schools'. CT measure (finalised SHU)
Jan 2014- **March 31st** 2015	Phase 1: Setup *Recruitment* **Inform schools as to whether trial or control between May 1st- 7th 2015** Cont'd.	• Recruitment of 100 2-form entry schools (115 to allow for attrition), 2 teachers per school, with 23 schools across 5 English hubs, (only one in London). • Recruitment events, one per hub • Set up online community in each hub • Schools will be randomized by SHU into 50 treatment and 50 control schools (**May 1st 2015 half-term**)
June 2015	Phase 2: Training 1	• 2 days training of treatment school teachers in each of 5 hubs, 2 teachers per school (one year 5 teacher/one ICT coordinator) trained in CT intervention. After SATS, with couple of weeks between training days.
Sept 2015 – Apr 2016	Phase 3: Delivery 1	• Treatment schools deliver the CT intervention to Year 5 pupils, coordinated by hubs. • Observation interviews to research causal mechanism.
October 2015	Phase 3: Design 2	• ScratchMaths intervention finalized and PD.
May 2016	Phase 4: Testing 1	• Testing using CT measure with Year 5s in treatment schools (SHU)
June 2016	Phase 4: Training 2	• 2 days training for treatment schools. 2 Year 6 teachers per school, in ScratchMaths intervention across the 5 hubs (again with a week or so between each day) • Training in CT in Control schools
Sept 2016- Apr 2017	Phase 5: Delivery 2	• Treatment schools deliver ScratchMaths intervention to Year 6 pupils (Y5 from last year) • Observation interviews to research causal mechanism • Control schools teach CT intervention to Year 5
May 2017	Phase 5: Testing 2	• Final post-test, KS2 maths in all treatment schools

References

Back, J Hirst, C; De Geest, E; Joubert, M and Sutherland, R. (2009). *Final report: Researching effective CPD in mathematics education (RECME)*. NCETM

Clements, D. (2000). From exercises and tasks to problems and projects unique contributions of computers to innovative mathematics education. *Journal of Mathematical Behavior,* 19, 9-47.

DfE (2013. *Computing programmes of study: Key stages 1 and 2. National curriculum in England*. DfE

Furber S., (2012). Shut down or restart? The way forward for computing in UK schools. The Royal Society.
Holmes, W.M. (2014) Using Propensity Scores in Quasi-Experimental Designs. Sage.

Lee, W C. (1990). *The effectiveness of computer -assisted instruction and computer programming in elementary and secondary mathematics: A meta-analysis* PHD dissertation, University of Massachusetts (January 1, 1990). *Electronic Doctoral Dissertations for UMass Amherst.* Paper AAI9022709.

McMMaster, K, Rague, B., & Anderson, N. (2010). *Integrating mathematical thinking, abstract thinking, and computational thinking.* Paper presented at ASEE/IEEE Frontiers in Education Conference, October 2010, Washington DC.

Monroy-Hernandez, A. & Resnick, M. (2008). Empowering kids to create and share programmable media. *Interactions* March and April 2008, 50-53.

Opfer, V. D. and Pedder, D., (2011). Conceptualizing teacher professional learning. *Review of Educational Research,* 81, 376-407.

Resnick, M. etal. (2009). Scratch: Programming for all. *Communications of the ACM,* 51(11), 60-67.

Stoll, L., Harris, A. & Handscome, G. (2012). *Great Professional Development that leads to great pedagogy: nine claims from the research*. NCTL.

Van Driel, J. et al. (2012). Current trends and missing links in studies on teacher professional development in science education: a review of design features and quality of research. *Studies in Science Education,* 48:2, 129-160.

Yoon, K. S. et al. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007-No.033). Washington, DC: Department of Education, Institute of Education Sciences, NCEERA