Statistical Analysis Plan for ScratchMaths

Sheffield Institute of Education, Sheffield Hallam University

INTERVENTION	**ScratchMaths**
DEVELOPER	Institute of Education/ London Knowledge Lab, University College London
EVALUATOR	Sheffield Institute of Education, Sheffield Hallam University
TRIAL REGISTRATION NUMBER	ISRCTN10189078 http://www.isrctn.com/ISRCTN10189078
TRIAL STATISTICIAN	Sean Demack
TRIAL CHIEF INVESTIGATOR	Mark Boylan
SAP AUTHOR	Sean Demack
SAP VERSION	2
SAP VERSION DATE	18/5/17
EEF DATE OF APPROVAL	
DEVELOPER DATE OF APPROVAL	

Table of contents

Table of contents 2
Introduction 3
Study design 4
Protocol changes 4
Randomisation 5
Calculation of sample size 6
Follow-up 7
Outcome measures 8
    Primary outcome 8
    Secondary outcomes 8
Analysis 8
    Primary intention-to-treat (ITT) analysis 8
    Imbalance at baseline 9
    Missing data 10
    On-treatment analysis 11
    Secondary outcome analyses 13
    Additional analyses 13
    Subgroup analyses 13
    Effect size calculation 14
Enrolment 15
Allocation 15
Follow-Up 15
Analysis 15
Analysis 15
Follow-Up 15
Enrolment 15
Allocation 15

Introduction

Scratch is a freely available programming environment developed for educational purposes at the Massachusetts Institute of Technology in the early 21st Century (Monroy-Hernandez & Resnick, 2008). The University College London Institute of Education (IoE) ScratchMaths team designed and developed an educational intervention using Scratch that aimed to increase attainment in Key Stage 2 mathematics - the intervention was known as ScratchMaths and ran for two academic years (2015/16 & 2016/17) and was preceded by an intervention pilot and development year (2014/15).

ScratchMaths was evaluated using a research design that centred on a clustered randomised controlled trial (CRT)

The evaluation aimed to answer a number of research questions, the key question in relation to the impact of ScratchMaths on the primary outcome (KS2 mathematics attainment) was

- What has been the effect of the intervention on the development of pupils' mathematical skills as measured by a randomised control trial?

The intervention provided two days of training in the use of Scratch programming and the ScratchMaths materials[1] for two primary school teachers in both years (Y5 teachers in year 1, Y6 teachers in year 2). In the first year, the training focused on computational thinking and Scratch Programming. In the second year, the training focused on the learning of mathematical thinking through engagement with specially designed Scratch-mathematics curriculum and tasks. This training was aligned with both the Primary Computing and Primary Mathematics National Curricula. PowerPoint presentations, video animations, and an explicit pedagogic framework were also made available and two further (optional) half days of training offered to Y5 teachers in 2015/16 and one half day with on-line webinar support to Y6 teachers in 2016/17.

The theory of change outlined in the ScratchMaths protocol hypothesised that the direct impact of the ScratchMaths intervention on overall KS2 mathematics attainment would be weak. The strength of impact on KS2 mathematics attainment was hypothesised to depend on three things; the development of computational thinking (which is taken to be related to mathematical thinking); on participating teachers responding to PD to make links between computational thinking and Scratch programming and mathematics; and on the breadth and depth of coverage of some KS2 mathematics curriculum content within ScratchMaths material.

An interim, secondary outcome that measured computational thinking was developed and piloted in 2014/15 and the resulting computational thinking test was administered to two participants at the end of the first year of the trial when pupils were in Y5 in July 2016.

A 3-level regression model will be constructed to measure the direct impact of ScratchMaths on the primary outcome KS2 mathematics attainment[2]. Follow on multilevel analyses will measure the impact of ScratchMaths on the secondary outcome (computational thinking) and the statistical relationship between the primary and secondary outcomes.

[1] Detailed ScratchMaths tasks, ScratchMaths files, offline 'unplugged' tasks and challenge & extension tasks

[2] The ScratchMaths Design and implementation team has critiqued aspects of the design of this CT measure.

Study design

111 primary schools were recruited to the trial and were located in seven geographical hub areas[3]. IoE recruited the schools using two basic criteria: that the schools delivered Key Stage 2 and had a two form entry. Subsequently, at IoE's request, due to concerns with recruitment, SHU relaxed this second criterion and in the randomisation spreadsheet, there are schools with between one and four form entry. The implication for schools with 3-form entry is that the school level data for Maths KS2 measure will include results for pupils whose teachers did not participate in professional development[4]. **However, prior to randomisation schools provided class lists who would be taught teachers who would be doing the ScratchMaths CPD. This is the participant sample for impact analysis at pupil level regardless of school size.**This will be an issue to consider alongside others in the Implementation and Process Evaluation related to participation of teachers in professional development, use by teachers of ScratchMaths materials and attrition. IoE's fidelity measure allows for cascade or in school PD as a substitute for attendance. In addition, in the impact analysis any significant patterns in relation to school size can be considered.

A clustered randomised controlled trial design was adopted with randomisation at the school level. In April 2015 110 of the 111 recruited schools were randomised into the intervention (or wave 1) or control (or wave 2) groups. A waitlist approach was adopted as detailed below:

Intervention (Wave 1) Schools (2,986 pupils in 55 schools)

- ScratchMaths Professional Development events for Y5 teachers in summer 2015 and Y6 teachers in summer 2016.
- Schools implementing ScratchMaths with Y5 classes in 2015/16 and with Y6 in 2016/17.

Control (Wave 2) Schools (3,246 pupils in 55 schools):

- ScratchMaths Professional Development events for Y5 teachers in summer 2016 and Y6 in summer 2017 (after trial end).
- Pupils / teachers in Y5 during 2015/16 and Y6 in 2016/17 represent the 'business as usual' control group.

KS1 NPD data for attainment, FSM status and gender was obtained in January 2016. This data was collected in 2013, prior to randomisation. NPD data for the primary outcome (KS2 mathematics attainment) will be requested in June 2017. Data for the secondary outcome (computational thinking) was collected in July 2016. This is an interim secondary outcome and will not be collected again at the end of 2017.

Note that the waitlist design in this case meant that Y5 pupils in the control schools received the intervention in the second year of the trial. There is, then, a risk of potential spill-over from the those Y5 teachers and classes to the control Y6 teachers and classes. Data investigating the possible spill-over will be collected as part of the implementation and process evaluation, (IPE) by survey of teachers.

Protocol changes

The SAP was formulated after revision to the protocol undertaken in 2016 to update in relation to implementation and other matters. These changes are detailed on the updated protocol. Further, in reviewing the protocol, it is important to note that the waitlist design does not mean that the control schools received the intervention after the end of the trial, but

[3] Blackburn, Bradford, North London, South London, Merseyside, Somerset/Devon and Staffordshire.
[4] The initial school randomization spreadsheet indicates: Intervention group: 4 1-form entry; 2 1.5 form entry and 6 3-form entry. Control group: 1 1- form entry; 8 3-form entry

rather with a time lag of one year (see above). On page 9 of the protocol[5] it was incorrectly stated that "55 pairs were allocated into the control group, being offered the Scratch programming intervention **after the trial has ended**". As stated above, in this trial the 55 control schools were offered ScratchMaths during the trial period in 2016/17 for Y5 pupils. Whilst pupils in Y6 should not directly receive the ScratchMaths intervention, there is a potential for some spill-over which will be investigated as part of the IPE.

A further change since revision of the protocol is that the ScratchMaths team requested a change to the fidelity measure for Y5 and also proposed a fidelity measure for Y6 - see below on on-treatment analysis.

Randomisation

A stratified approach was adopted for the school level randomisation in April 2015. As detailed in the protocol, a logistic regression model was used to generate school level predicted probability or “propensity” scores based on the schools’ 2013/14 KS2 attainment outcome variable[6] and seven explanatory variables[7]. 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 geographical hub areas. This was not the case for three: an odd number of schools were recruited in both of the London hubs and in the Somerset hub. The two London hubs (north and south) were merged into a single hub with an even number of schools prior to pairing schools within the hub 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 were paired and randomised to the intervention or control group. The Somerset school that was excluded from the trial will still be offered ScratchMaths as part of the waitlist design but we will not be using any data from this school.

In all, 55 schools were randomly selected to receive the ScratchMaths intervention and their 55 pairs were allocated into the control group.

In 2014/15, prior to randomisation, recruited schools were requested to provide lists of all pupils in Y4 and the name of their class and teacher[8]. Following randomisation in April 2015, 2,986 pupils were located in 97 classes in the 55 intervention schools and 3,246 pupils located in 110 classes in the 55 control schools.

The propensity-score-paired-school-stratification approach to randomisation brings three key advantages:

- It enables a large number of variables to be drawn on to stratify
- Stratification variables can be finely grained (scale)
- Flexibility and robustness for follow on analyses (such as on-treatment or sensitivity analyses)

[5] See https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Round_6-_Scratch_maths_amended.pdf

[6] 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).

[7] Explanatory variables - KS1 attainment, KS1 to KS2 progress in mathematics, school size, gender balance, %FSM, %EAL, %SEN.

[8] Pupil name, date of birth, Unique Pupil Number (UPN), teacher and class name were requested.

The first two of these are self-explanatory. Our propensity-score-paired-school-stratification approach drew on seven scale variables and then created 55 pairs of schools[9]. Within each pair, one school was randomly selected into the intervention (wave 1) and the other to the control (wave 2) group.

The third advantage is most clearly illustrated with respect to on-treatment analyses.

An on-treatment analysis might proceed from an analysis that examined whether fidelity to an intervention was statistically associated with a trial outcome. If an association between fidelity and an outcome is found, a subsample of the intervention group might be identified as being 'on treatment' by reaching a specified level of 'fidelity'.

An on-treatment impact analysis might then compare the outcome(s) for this restricted 'on treatment' intervention group subsample with a control group. With minimisation or more standard stratification approaches to randomisation, the original complete control group would usually be used for this comparison. This brings an increased risk of imbalance between the restricted 'on-treatment' intervention subsample and the original control sample.

Our propensity-score-paired-school-stratification approach brings a flexibility that will limit this risk of imbalance. Once an 'on-treatment' intervention subsample of schools is identified, the control group can similarly be restricted to just include the matched pairs for each 'on-treatment' school. For example, if 30 of the 55 intervention schools involved in the trial are identified as 'on treatment', these 30 intervention schools could be compared with their 30 matched control schools rather than the entire 55 control school sample.

A similar approach could be taken for sensitivity analyses. For example, if data is not available for entire schools (e.g. with primary data collection- as was the case here for the interim secondary outcome; a computational thinking test in 2016).

A reduced sample would result in reduced statistical sensitivity but given that the 55 samples of 2 schools were determined prior to randomisation (using propensity scores), the sample balance advantages are provided without additionally compromising the RCT design.

Calculation of sample size

As specified in the protocol, a 3-level Clustered Randomised Controlled Trial design was adopted for this evaluation (pupils clustered into classes clustered into schools). Randomisation took place at the school level and the outcome variables are both at the individual pupil level. A class level was also included to reflect the structural reality of the data and to acknowledge the widespread use of setting within primary schools for KS2 mathematics. The term ' class level' refers to the class in which a pupil is located being accounted for in the multilevel analysis and so takes account of potential effects of clustering on outcomes.

The power calculations were undertaken using the Optimal Design Software[10]. Table 1 below summarises the estimated Minimum Detectable Effect Sizes (MDES) for the primary and secondary outcomes based on baseline numbers of pupils, classes and schools. The baseline numbers were very similar to those estimated in the protocol and the resulting MDES estimates have not changed. Table 2 summarises the sample size at baseline for the ScratchMaths CRT.

[9] This is set out in detail in the trial protocol https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Round_6-_Scratch_maths_amended.pdf

[10] Raudenbush, S. W., et al. (2011). Optimal Design Software for Multi-level and Longitudinal Research (Version 3.01) [Software]. Available from www.wtgrantfoundation.org

It was estimated that 13% of the variation in KS2 maths attainment will lie at the school level and 7% will lie at the class level[11].

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 1: Minimum Detectable Effect Size (MDES) for planned analyses for Scratch programming clustered RCT based on baseline data

- **Number of schools:** **110**
- **Number of classes:** **207**
- **Number of pupils:** **6,232**

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

Table 2: **ScratchMaths CRT: Summary of school, class & pupil numbers at baseline**

	Schools n=	Mean pupils per school	Classes n=	Mean pupils per class	Pupils n=
Wave 1 (intervention)	55	54.3	97	30.8	2,986
Wave 2 (control)	55	59.0	110	29.5	3,246
All	110	56.7	207	30.1	6,232

Follow-up

Please see the Appendix for a CONSORT flow diagram for the primary and secondary outcomes. At the time of writing, full details are available for the secondary outcome (computational thinking, collected in July 2016) and so this flow-diagram will not need to be updated for the final report. NPD data for the primary outcome (KS2 mathematics attainment) will be requested in June 2017 and so the flow-diagram may need to be updated to note any missing KS2 mathematics detail.

As can be seen from the flow-diagram, the computational thinking secondary outcome measure has a sizable issue with missing data. Much of this relates to schools dropping out of the intervention or not being able to administer the CT test. No CT test data was obtained from 15 intervention schools (1,162 pupils) and 14 control schools (1,108 pupils). This

[11] 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).

missing data combines with six pupil/parents opt outs to result in CT data for 40 intervention schools (1,820 pupils) and 41 control schools (2,136 pupils). Much of the loss of data relates to the 29 schools where no CT test was administered. Amongst the 81 schools where the CT test took place, the pupil response was very good for both the intervention (85% response) and control (88%) group pupils.

Because the primary outcome draws on NPD data, the analyses will not suffer from the same degree of missing data.

As noted below, the impact of missing data on balance between the intervention and control groups will be examined prior to the main impact analyses.

Outcome measures

Primary outcome

The primary outcome is KS2 mathematics attainment in May 2017[12] as measured by the NPD 'KS2_MATMRK' variable. Follow-on analyses will examine impact within the three KS2 mathematics test papers[13]. NPD data for KS1 attainment for the cohort under study was attached to pupil lists provided prior to randomisation and provided in January 2016. For the KS2 data, the same pre-randomisation pupil lists will be used for attaching the NPD data that will be requested in June 2017.

Secondary outcomes

The secondary outcome is computational thinking (CT) based upon pupil scores for a CT test developed and piloted by us in 2015 and administered to trial participants in July 2016.

Analysis

Primary intention-to-treat (ITT) analysis

An ITT analysis will only be feasible for the primary outcome (KS2 mathematics attainment). CT data was not obtained from 29 schools including 11 intervention schools known to have withdrawn from engagement in ScratchMaths professional development or use of materials [14]. Therefore, interpretation of analyses that involve the CT measure will need to be treated with greater caution.

Impact on KS2 mathematics attainment will be examined using a multilevel ITT analysis with three levels (pupils clustered into classes clustered into schools). As specified in the protocol, the model will be constructed in three stages. 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 KS1 maths covariates at both pupil and school level[15]. It is the findings from this second stage that will be used to assess 'impact' of the Scratch Programming intervention

[12] As part of a consideration of limitations of the trial the relationship between the 2017 SATS content and the ScratchMaths materials will be discussed.
[13] Specific NPD variables - KS2_MATARITHMRK (paper 1, arithmetic); KS2_MATPAPER2MRK (paper 2, reasoning) & KS2_MATPAPER3MRK (paper 3, reasoning).
[14] The total of 11 is based on information from the ScratchMaths team or provided by schools when asked to undertake the CT test. The situation of a further 4 schools is ambiguous.
[15] KS1 mathematics attainment will be included as a pupil level (KS1_MATPOINTS) and an aggregated version will be included at the school level.

on the primary outcome. A third model stage will include all of the school level variables that were included as explanatory variables[16] to generate the propensity-scores used for randomisation and dummy variables to identify school pairs within geographical hubs. This third and final stage of the primary ITT analyses are sensitivity analysis.

All scale / continuous explanatory variables will be centred around their mean.

STATA will be used for the main impact analyses, an example of the STATA SYNTAX is shown below.

Example of STATA SYNTAX for impact analyses

* **Stage 1**
.mixed **OUTCOME Group dummy** || School level Identifier: || class level identifier:
estat icc

* **Stage 2**
.mixed **OUTCOME Group dummy Pupil-level KS1 Maths (centred) School level aggregated KS1 Maths (centred)** || School level Identifier: || class level identifier:
estat icc

* **Stage 3**
.mixed **OUTCOME Group dummy Pupil-level KS1 Maths (centred) School level aggregated KS1 Maths (centred) [school level variables used for randomisation - centred] [Dummy variables identifying school pairs within geographical hub region]**||
School level Identifier: || class level identifier:
estat icc

Imbalance at baseline

Imbalance at baseline has been examined at the pupil level using the KS1 data obtained in January 2016. Table 3 summarises balance in terms of KS1 maths attainment, gender & FSM. Balance at the school level is shown in the protocol.

Table 3: Comparison of intervention & control groups at baseline

	KS1 Maths Points Score	% FSM	% Female
	Mean (sd)		
Wave 1 (intervention)	16.1 (3.44)	28.7	49.6
Wave 2 (control)	16.0 (3.44)	28.3	49.8
All	16.0 (3.44)	28.5	49.7

Balance at both school and pupil level at baseline was observed to be good / excellent.

[16] School level variables - KS1 attainment, KS1 to KS2 progress in mathematics, school size, gender balance, %FSM, %EAL, %SEN.

Missing data

A comparison of the baseline balance with what is observed for the samples with primary / secondary outcome data will be undertaken. No imputation is planned but if evidence of imbalance is observed in terms of %FSM and/or %Female, a model that includes the FSM and/or gender variable will be constructed to examine how the imbalance might influence estimating the impact of ScratchMaths (this would represent additional sensitivity analyses)

On-treatment analysis

IoE have recorded the status of 15 schools (January 17) recruited to the intervention who have', 'withdrawn' or 'may withdraw'. Withdrawn means the school has contacted IoE and said they are not using the materials. 'May withdraw' means they haven't been in contact with IoE or may have not attended PD, and so there is a doubt about whether materials are being used. There are a further 40 schools who are assumed to be implementing the project to some extent.

In addition, a teacher level classification is recorded of being 'active' or 'non-active', the latter being used, for example, if there has been a change of teacher implementing the project.

Taken together, these designations provide a simple immediate criteria for sensitivity analysis.

As part of the implementation and process evaluation, data on implementation will be reported including patterns of use of curriculum materials and teacher attendance at PD events[17]. In addition this data will form the basis for an analysis of fidelity using the school level measures set out in Table 4:

Table 4: Defining high, medium and low fidelity to the ScratchMaths intervention in terms of attendance of PD, technology, coverage, time and progression.

	High	**Medium**	**Low**
1. Professional Development **Data source:** **SloE PD attendance data**	Y5 teacher attended at least two days of PD or equivalent (defined as any combination of Summer 2015 PD days or half-day optional PD or substantial in-school PD via cascade/co-planning with a teacher who attended PD) Y6 teacher attended at least two days of PD or equivalent (defined as any combination of Summer 2016 PD days or half-day optional PD or substantial in-school PD via cascade/co-planning with Y5 teacher or another teacher who attended PD)	Y5 teacher attended at least one day of PD or equivalent (defined as any combination of Summer 2015 PD days or half-day optional PD or substantial in-school PD via cascade/co-planning with a teacher who attended PD) Y6 teacher attended at least one day of PD or equivalent (defined as any combination of Summer 2016 PD days or half-day optional PD or substantial in-school PD via cascade/co-planning with Y5 teacher or another teacher who attended PD)	Y5 teacher had some form of limited PD with a teacher who had attended PD, their SMLC or a member of the SM team. Y6 teacher had some form of limited PD with a teacher who attended PD or taught Y5 SM, their SMLC or a member of the SM team.
2. Technology **Data source:** **Teacher Survey**	Computers running Scratch 2.0 offline or adequate internet access Minimum 2:1 pupil to computer ratio	Computers running Scratch 2.0 offline or adequate internet access Minimum 2:1 pupil to computer ratio	Computers running Scratch 2.0 offline or adequate internet access Minimum 3:1 pupil to computer ratio
3. Coverage **Data source:** **Teacher Surveys**	Pupils taught at least some of the core activities across 5 different modules	Pupils taught at least some of the core activities from across 4 different modules	Pupils taught at least some of the core activities from across 3 different modules
4. Time **Data source:**	Time spent on teaching is at 20+ hours in Y5 and at least 12+ hours in Y6.	Time spent on teaching is at least 12+ hours per year.	Time spent on teaching is less than 12 hours per year.

[17] This will be important when considering issues of fidelity, scalability and limitations of the evaluation in the final report.

Teacher Surveys			
5.Progression **Data source:** **Teacher Survey**	The order of modules and order of activities are mostly followed in general	The order of modules and order of investigations are mostly followed in general	The order of modules are mostly followed in general

The statistical relationship between fidelity to the ScratchMaths intervention and the primary and secondary outcomes will be explored descriptively. These analyses will explore fidelity across the five components shown in Table 4. An overall fidelity measure will then be derived that will be used to identify schools with high, medium or low fidelity to ScratchMaths as shown in Table 4.

If these analyses reveal that fidelity to ScratchMaths is correlated with either outcome, an exploratory follow-on on-treatment analysis will be undertaken. This will only include intervention schools with high or medium/high fidelity. For the on-treatment analysis, as outlined above, the control sample will be constructed by returning to the propensity scores used at randomisation to select the subsample of control schools that were matched to the intervention schools with high or medium/high fidelity to ScratchMaths. The on-treatment analysis approach will take the same multilevel approach used for the main ITT impact analyses.

Secondary outcome analyses

As noted above, an ITT approach for the secondary outcome (computational thinking) is not possible due to issues of missing data. Analysis of the secondary outcome will adopt the same three-level and three stage multilevel analysis taken with the primary outcome; First, an outcome only analysis; Second, the model will include KS1 maths covariates at pupil and school levels; Third, the model will include all of the school level variables that were used as explanatory variables to generate the propensity-scores at randomisation and dummy variables to identify school pairs within geographical hubs. The second stage model will be used to assess the impact of ScratchMaths on computational thinking whilst the third stage represents sensitivity analyses.

Additional analyses

Follow-on analyses relating to the primary outcome will be undertaken. Specifically, these analyses will focus on attainment within the three KS2 maths test papers. These analyses will adopt the same ITT approach used for the main primary analyses and exactly the same three-level and three stage multilevel analytical approach.

Further follow-on exploratory analyses will explore the relationship between KS1 mathematics, computational thinking and KS2 mathematics attainment. The specific focus of these analyses will be to examine whether there is an interaction between the ScratchMaths intervention and computational thinking in relation to KS2 mathematics attainment. If an interaction is found to be statistically significant, this will suggest that the impact of ScratchMaths on KS2 mathematics in Y6 attainment is dependent on the level of computational thinking that a pupil attains in Y5.

A three-level multilevel analysis will be undertaken with KS2 mathematics as the outcome and adopting two model stages; First, a main effects model that includes the dummy variable that identifies whether a pupil is in intervention or control group; KS1 maths attainment, computational thinking and KS2 attainment. Second, an interaction between the group identifier dummy and computational thinking will be included. If the interaction is observed to be statistically significant, a follow-on analysis will be undertaken. The follow-on analysis will split the sample into two parts based upon the computational thinking variable (i.e. a higher and lower CT attainment group around the median CT value). Impact analyses will then be undertaken on both upper and lower CT score samples to assess whether ScratchMaths had differential impact for pupils with different levels of CT.

Subgroup analyses

Exact detail on the planned subgroup analyses are dependent on findings from including interaction terms into the primary and secondary outcome models.

Three interaction terms will be included:

Group identifier * KS1 maths attainment - to examine whether the impact of ScratchMaths is different for pupils with differing levels of KS1 maths attainment.

Group identifier * gender - to examine whether the impact of ScratchMaths is different for males and females.

Group identifier * FSM - to examine whether the impact of ScratchMaths is different for pupils classed as FSM compared with pupils classed as not FSM[18].

For gender and KS1 attainment, if the interaction term is found to be statistically significant, follow-on subgroup analyses will be undertaken (e.g. separate impact analyses for males / females; higher / lower KS1 maths attainment.

For FSM, the subgroup analyses will be undertake regardless of whether the Group*FSM interaction term is statistically significant as requested by EEF.

Effect size calculation

The impact of ScratchMaths will be measured using the Hedges g effect size statistic based on the formula shown below.

$$ES = \frac{(T - C)_{adjusted}}{\sqrt{\delta_s^2 + \delta_c^2 + \delta_p^2}}$$

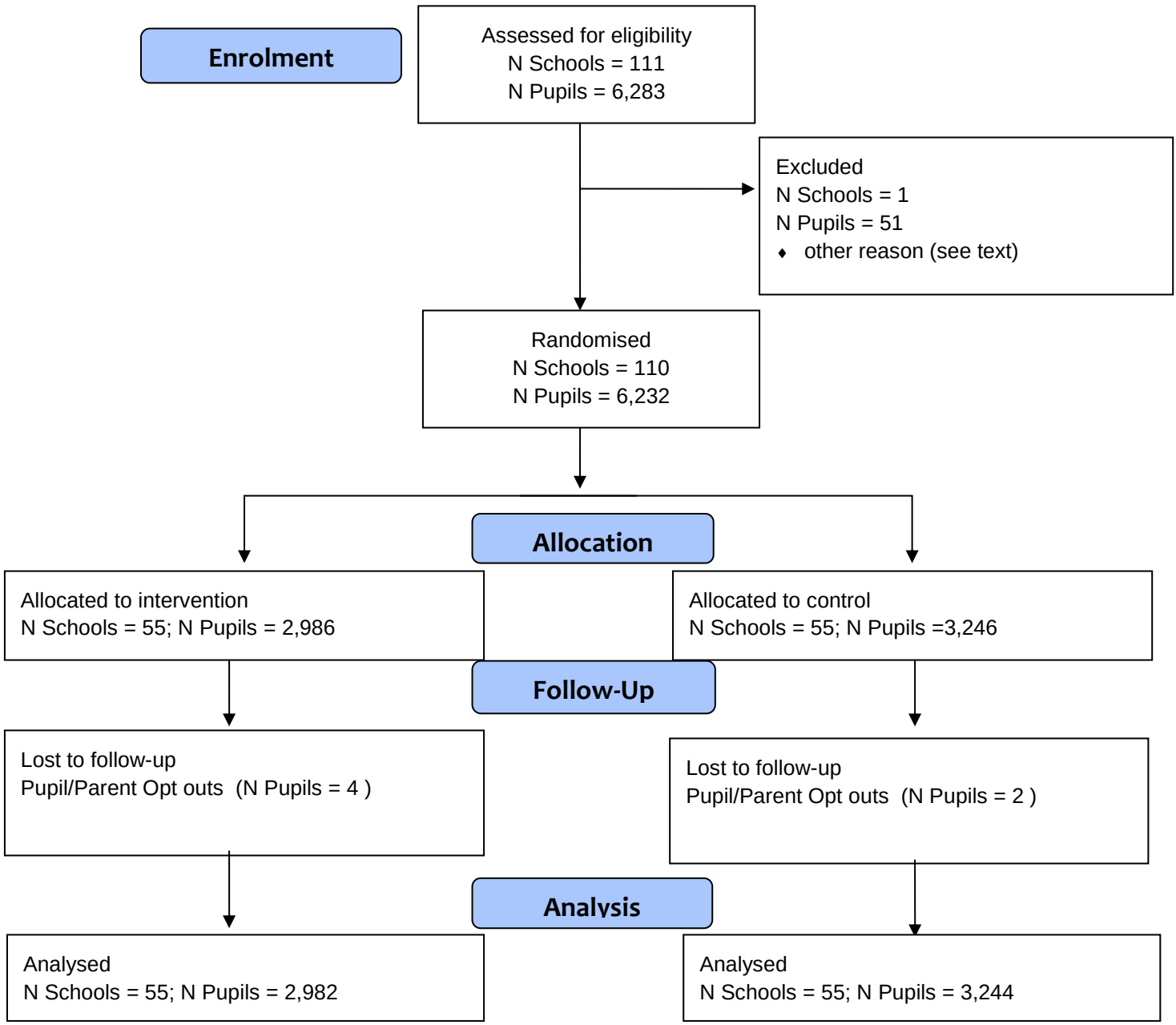
Where δ_s^2 is the school level variance, δ_c^2 is the class level variance and δ_p^2 is the pupil level variance and $(T - C)_{adjusted}$ is the coefficient estimate for the group identifier dummy variable from the multilevel model.

The upper and lower confidence intervals for the coefficient estimate for the group identifier dummy variable will also be standardised into Hedges g effect size statistics.

Effect sizes and confidence intervals will be calculated for all model stages[19].

[18] The 'EverFSM' variable will be used here.
[19] This will include the sensitivity analyses that include dummy variables identifying school pairs within hubs.

ScratchMaths Interim* CONSORT CRT Flow Diagram:
Primary Outcome
(KS2 Maths Attainment in Y6 in May 2016)



*** NOTE -** *This flow diagram may need to be updated if any KS2 mathematics data is missing.*

ScratchMaths CONSORT CRT Flow Diagram: Interim / Secondary Outcome (Computational Thinking in Y5 in July 2015)