INTERVENTION	**1stClas@Number**
DEVELOPER	Nick Dowrick, Edge Hill University
EVALUATOR	Terezinha Nunes, Rossana Barros, Maria Evangelou, Steve Strand, Sandra Mathers, University of Oxford
TRIAL REGISTRATION NUMBER	Protocol serial number - 37482473.1 ISRCTN15932880 DOI 10.1186
TRIAL STATISTICAL ANALYSIS BY	Rossana Barros, Terezinha Nunes, Steve Strand
TRIAL CHIEF INVESTIGATOR	Terezinha Nunes
SAP AUTHORS	Terezinha Nunes, Rossana Barros, Steve Strand
SAP VERSION	1
SAP VERSION DATE	24 May 2017
EEF DATE OF APPROVAL	15 August 2017
DEVELOPER DATE OF APPROVAL	17 July 2017

Protocol and SAP changes

One small change was made to the protocol, regarding the number of geographical areas for blocks in the randomisation process; it is pointed out in this document.

Table of contents

Contents

Introduction

This evaluation aims to test the impact on student outcomes of the programme *1st Class@Number,* which is an intervention designed by Edge Hill University. The programme aims to support those children who in Year 2 are struggling with the mathematics curriculum. The intervention is delivered by especially trained teacher assistants (TAs) to small groups of four children who, at the start of Year 2, are judged by their teachers as having difficulties accessing the Year 2 curriculum and may benefit from taking part in 1st Class@Number. TAs receive six sessions of professional development and deliver five topics in six half-hour lessons each (a total of 30 half-hour sessions), normally three times a week over approximately 10 weeks. These sessions are in addition to usual, daily classes of mathematics.

The current widespread use of 1st Class@Number and the efficacy of a similar intervention, Numbers Count, which is delivered by highly trained teachers rather than by TAs, indicate that 1st Class@Number should be evaluated in an effectiveness trial.

The primary research question in the project is:

- Do the children identified by their teachers as struggling with mathematics at the start of Year 2 who participate in the 1st Class@Number intervention perform significantly better in the Quantitative Reasoning Test than children identified by their teachers as struggling with mathematics who do not participate in the intervention?

The statistical analyses will test whether the children who participated in the 1st Class@Number intervention differ significantly from the control group, after controlling for pre-test results. We will estimate the effect size and analyse whether the conclusions are robust by considering other aspects of the trial, such as fidelity of implementation and risks to validity such missing data (percent; not at random loss of participants).

The secondary research questions are:

- Do the children identified by their teachers as struggling with mathematics at the start of Year 2 who participate in the 1st Class@Number intervention perform better in the Key Stage 1 Maths than the children identified by their teachers as struggling with mathematics who do not participate in the intervention?

Key Stage 1 Maths is a secondary outcome measure; therefore analyses with this measure are treated as a secondary. The analyses will consider whether the effect, if any, could be due to chance, the effect size, and the robustness of the conclusions.

- Do children entitled to free school meals (FSM) benefit equally as other children from the 1st Class@Number intervention as assessed by the Quantitative Reasoning Test?

- Do children entitled to free school meals (FSM) benefit equally as other children from the 1st Class@Number intervention as assessed by Key Stage 1 Maths?

The statistical analyses will investigate whether there is a main effect of membership in the group of children eligible for free school meals (using the EVER6 variable from the National Pupil Database, which indicates eligibility for free school meals in the last six years) and whether there is an interaction between membership in this group and membership in the intervention vs control group. If the interaction is significant, specific group comparisons will also be used.

- Is the 1st Class@Number intervention as effective for girls and boys as assessed by the Quantitative Reasoning Test?

- Is the 1st Class@Number intervention as effective for girls and boys as assessed by Key Stage Maths?

The statistical analyses will investigate whether there is a main effect of gender and whether there is an interaction between gender and membership in the intervention vs control group.

Study design

Participants

The study was located in schools in the South and West Yorkshire. All schools that had not previously participated in 1stClass@Number were eligible to participate. A briefing meeting was scheduled with schools to inform them of the design and requirements of treatment implementation; two dates, each at a different location, were offered and schools were free to choose which briefing meeting to attend. Schools that were not represented at the meeting were not included in the sample. 138 schools were recruited for the briefing meeting; 5 schools withdrew before the pre-test and randomisation; 133 schools were randomised; 67 schools were allocated to the intervention and 66 schools to the control group. At the point of post-testing, 3 control schools have withdrawn from the trial and 1 intervention school had interrupted the intervention but agreed to participate in the post-test.

As the intervention is designed for small group work, the intervention team instructed teachers during the briefing meeting on how to proceed for the selection of participants. Pupils selected for the intervention should be those who are perceived by the teachers as struggling with maths. Teachers could use information from tests as well as their own knowledge of the pupils. Teachers were also advised to nominate for the same group children who can "gel" and learn together and who were not participating in another intervention.

A pre-test was delivered by teachers at the end of Year 1 (see design section). At the beginning of Year 2, the evaluation team informed the teachers of the percentile that described the performance of each child in the class using norms based on the overall sample of 5,353 children from the same year group and in the same region. Teachers then nominated the children for participation using personal judgement and results from the pre-test. Children who had not taken the test at the end of Year 1 were not eligible for nomination.

- Description of trial design

This is a two-arm trial, with an intervention and an unseen, business-as-usual control group. The design is intention to treat and included a pre-test and a post-test. Randomisation was at school level.

The unit of randomisation was the school in order to avoid contamination between groups. The control schools were asked to take a business as usual approach to supporting pupils who struggle with mathematics.

The implementation of the intervention was outside the time dedicated to regular mathematics lessons in the school. Thus, it is an intervention that includes extra time dedicated to mathematics learning. Sessions were scheduled at different times on different days to avoid pupils missing often the same lessons.

- Sample size

A total of 532 children were nominated for participation, 4 per school. If nominated children changed schools and the TA included a new child in the group in order to maintain the size

of the group constant, which is necessary for the purposes of the intervention implementation, the late-entering child will not be included in the analysis. Reasons for schools dropping out of the programme were collected during the implementation and will be analysed in the evaluation of patterns of missing data.

- Description of trial arms

The intervention

1stClass@Number is an intervention delivered by TAs, with the support of a Link Teacher, to small groups of 4 children identified by their teachers as children who could benefit from the intervention (see criteria presented earlier on). TAs are trained by advisors especially trained by the 1stClass@Number team. Advisors are provided with a set of slides and a script for the six training sessions, which are roughly matched with the delivery of the five topics in the intervention: all about number; exploring place value; addition and subtraction 1; addition and subtraction 2; towards multiplication and division. The final session is an opportunity to reflect on the work done during programme delivery.

During training sessions, the advisors use PowerPoint slides prepared by the intervention team, which are accompanied by notes to be used during the presentation. The training sessions are clearly described and highly scripted, which enable a high level of consistency across trainers in training sessions. The sessions typically involve discussion of previous sessions and detailed preparation of the subsequent one. TAs have the opportunity to display their pupils' work, comment on implementation, and ask questions. Some scripting for how to implement specific activities is included in the training sessions. TAs receive materials that clearly describe each of the activities they will be implementing, but there is room for considering children's answers and how to approach errors. There is explicit guidance on how to assess children at the start of a new topic and how to adjust the activities to the pupils' performance in this initial assessment.

The 1stClass@Number team advises schools on the selection of TAs to implement the intervention. The advice is that TAs should:

- have successful experience of supporting children's mathematics
- be able to engage fully in training sessions that include the programme's procedures and the mathematics curriculum
- be able to make independent decisions while planning for and teaching children, based on an understanding of their needs and with the support of a Link Teacher**.**

The intervention team requires the school to nominate a Link Teacher to support the TA's work. Link Teachers were also nominated prior to randomisation. Guidance in the identification of the Link Teacher was also provided. Edge Hill recommended that the Link Teacher should be someone who would:

- provide the strategic leadership for the programme
- be suitably experienced to give support with the teaching of mathematics
- have time to support the teaching assistant and liaise with the class teacher(s), senior managers, SENCo, and other professionals as necessary
- attend two half-day training sessions with the teaching assistant.

Process evaluation data are being collected about the leadership and support provided by the Link Teacher, TAs' attendance at training days, and children's attendance at lessons.

The control schools

A representative of the all schools, including control schools, attended a briefing meeting. They all received training to implement the pre-tests. All pre-tests were administered and teachers from all schools received the same information about the children in their classes (i.e. control and intervention schools), which they could use to nominate the children. Randomisation took place after the children had been nominated and schools were informed of their group membership.

After being informed that they had been assigned to the control group, schools could choose to implement one alternative intervention produced by the intervention team, either on numeracy or literacy, which was appropriate for older children. The nominated children did not receive this alternative intervention.

Aside from this involvement with the intervention team, the schools were asked to do what they would normally do with Year 2 children and carry on as usual. 13 schools opted out of the use of the interventions for older children but agreed to allow post-testing of their children; these are not considered drop-out schools.

- Number and timing of measurement points

Pupils participated in a pre-test, which took place in June 2016, when the children were at the end of Year 1. They will be post-tested in a parallel form of the assessment during June 2017. The pupils were assessed in the Key Stage 1 maths in May 2017.

Randomisation

Randomisation took place after all the children had been assessed at baseline at the end of Year 1, whose information had been passed on to teachers in Year 2, and who had been nominated. Teaching Assistants (TAs) who would be implementing the programme were also nominated prior to randomisation, apart from one school where the TA who was expected to deliver the intervention fell ill and the school was still assessing whether to nominate the TA's substitute or to wait for the TA's return. Randomisation occurred in September 2016, when the children were already in Year 2.

The unit of randomisation was the school, in order to prevent contamination between the groups. Schools were clustered in five geographic areas[1], which were identified by the intervention team for delivery of the training to TAs; six blocks were created for randomisation, splitting the schools within one location (Leeds) in two groups, in order to keep in line with the training offered by the intervention team. The number of schools in each block is presented in Table 1.

Schools in each block were split in two blocks using the median for the number of children eligible for pupil premium from the database 2016-2017[2], resulting in 12 blocks. Random numbers were generated for all schools using SPSS. Schools were ordered by these random numbers in ascending order within their block. Each block included between 8 and 13 schools. The schools allocated the highest random numbers in each block were allocated to the intervention group until half of the schools had an allocation; those with the lowest random numbers were allocated to the control group.

The SPSS syntax used for this randomisation process was:

Syntax:
COMPUTE random=RV.UNIFORM(1,2).

[1] The protocol indicated four geographical areas but the intervention team identified five after recruitment.
[2] Source: (https://www.gov.uk/government/publications/pupil-premium-conditions-of-grant-2016-to-2017)

EXECUTE.

SORT CASES BY block(A) random(A).

Table 1. Number of schools by geographical blocks and by percentage of children eligible for pupil premium

Geographical Block	Number of Schools	Blocks defined by the median number of children eligible for pupil premium per school			
		Blocks with lower % of pupil premium	Number of schools	Blocks with higher % of pupil premium	Number of schools
Doncaster	24	Block 1 (11.28%)	12	Block 2 (38.23%)	12
Huddersfield	17	Block 3 (4.95%)	8	Block 4 (36.63%)	9
Leeds A	24	Block 5 (11.09%)	12	Block 6 (42.00%)	12
Leeds B	24	Block 7 (8.68%)	12	Block 8 (41.44%)	12
Sheffield	18	Block 9 (14.24%)	9	Block 10 (53.13%)	9
Wakefield	26	Block 11 (11.60%)	13	Block 12 (32.61%)	13
total	133				

Calculation of sample size

As specified in the protocol, we used Optimal Design software and made the following assumptions: (i) pupil outcomes measured at pre-test and at post-test have a correlation of r=0.70 at pupil level; (ii) a within school sample of 4 pupils per school; (iii) an intra-class correlation coefficient of 0.15; (iv), power of 0.80, alpha of 0.05 and a 2 tailed significance test. A sample of 60 schools per trial arm (120 schools in total) would yield approximately 480 children in total giving 80% power to observe an effect size of 0.22 SD. It was considered advisable to recruit 130 schools to allow for drop-outs after randomisation; 4 control schools dropped out after randomisation.

Measures

Pre-test

All children in the schools recruited for the project were assessed in the baseline measure prior to randomisation at the end of Year 1, subject to parental consent. The pre-test was delivered by teachers to all the children in their classes. Teachers were trained to implement the pre-test by the evaluation team during the briefing meeting. The test has a standardised presentation format for whole class administration. The questions are presented with support of drawings; no reading is required of children. The drawings are projected on a screen in front of the class and reproduced in the children's answer books. Teachers read the instructions for each item two times, with a pause between the two readings. Children write their answers on the answer books. Marking was done by the evaluation team using

standardised instructions. Marking is objective (i.e. it is verified whether the child wrote the correct number or marked the correct alternative), but if an assessor was in doubt about what a child had written (e.g. whether the number was a 2 or a mirror image of a 5, which is not uncommon at this age level), a second assessor verified the item. Marked tests were machine read for the creation of the data base.

In a previous study, this baseline measure showed a correlation of .7 with the outcome measure, when these were separated by a 12-months interval (Nunes, Bryant, Evans, & Barros, 2015), which is substantive enough for a covariate.

Outcome measures

The primary outcome measure chosen as post-test for the trial is the Quantitative Reasoning Test and will be administered by testers blinded to the schools' group membership at the end of Year 2. The post-test is a parallel form of the baseline measure.

The measure is delivered to whole classes. The form used for this trial includes 12 items that are identical to items used in the pre-test and 8 that are different and more difficult than those removed from the pre-test. Teachers did not keep copies of the pre-test and were not aware that some items at post-test would be the same. This combination of identical and new items allows for measuring progress from pre- to post-test while at the same time increasing the level of difficulty of the assessment at post-test.

A secondary measure will be used, the Key Stage 1 Maths, which is administered by the pupils' own teachers. Thus testers cannot be blinded and would know whether the pupils had participated in the 1st Class@Number intervention. These tests are not free from bias, but are familiar tests to schools. They can have a significant impact on pupils' lives as teachers sometimes use the results for streaming pupils by ability levels.

Key Stage 1 Maths results will be obtained through the NPD; these records contain only an ordinal scale of measurement with three points: Working Towards Expected Standard (WTS), Working at the Expected Standard (EXS) or Working at Greater Depth within the Expected Standard (GDS); there may be a small number working lower than WTS, at the Pre Key Stage foundation (PKF). For this measure, an ordinal regression (if the proportional odds assumption is met) or multinomial regression (if the proportional odds assumption is not met) will be used for the KS1 maths outcome.

Preliminary analyses

A total of 5,353 pupils participated in the pre-test. The mean number of correct responses (out of 20 items) for the whole sample was 9.41 (SD=4.59). The distribution was normal. Cronbach's alpha for the scale was .83. The percentage of pupils passing an item varied between 19% and 75%, which suggests an appropriate level of difficulty for the sample.

There was huge variation between schools; the lowest mean score by school was 3.64 and the highest was 13.41. The maximum score by school also varied significantly; the lowest maximum score by school was 8 and the highest was 20. The standard deviation by school also varied considerably, from 1.83 to 5.6. A one-way ANOVA showed that there was a significant effect of school (treated as nominal scale) on pupils' pre-test scores with this whole sample.

From this larger sample, 532 pupils were nominated for the intervention. The pupils nominated for the intervention were those judged by the teachers to be struggling with mathematics, so the mean for the nominated pupils was considerably lower than that for the reference group pre-tested: the mean was 5.02 and the SD was 2.42. Figure 1 displays the distribution of scores at pre-test for the sample of nominated pupils.

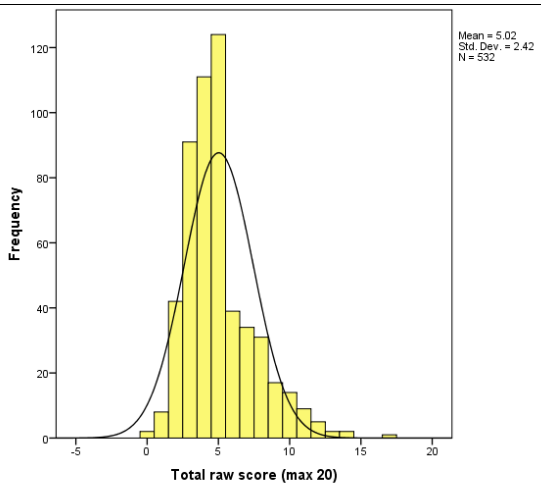

Figure 1. Distribution of scores in the pre-test for the pupils nominated for the 1stClass@Number intervention.

The mean for the control group was 5.15 (SD=2.49) and for the intervention group, 4.9 (SD=2.35). These means did not differ significantly (t=1.49; p=.25) and neither did the variance.

The pupils in the overall sample varied in age between approximately five and a half and seven and a half years; the mean age was approximately 6 years 4 months. The ages of the nominated pupils did not differ significantly between groups: the mean age of the pupils in the control group was 6.20 years and of the pupils in the intervention group was 6.23 years.

There were no significant differences between the groups at pre-test, but there was considerable variation between the pupils. Teachers were asked to identify pupils who were struggling with maths and who would work well together; the information about the pupils' percentile in the pre-test was not the only information that the teachers could take into account. An analysis of the variation between nominated pupils at pre-test showed that their scores varied between 0 and 17; 31 pupils scored above the mean for the population that took the pre-test. The level of difficulty of the post-test is considerably higher than that of the pre-test; there is no cause for concern that the pupils scoring above the mean for the population in the pre-test might not demonstrate the effects of the intervention due to ceiling effects at post-test.

We tested whether the variation in the nominated pupils' scores could be attributed to the effect of school membership, but a one-way ANOVA showed that the effect of school (treated as a nominal variable) on pre-test outcomes for the sample of nominated pupils was not significant.

In summary, no imbalance was detected at baseline between the trial groups and the effect of school membership on the pre-test scores of the nominated children was not significant. The final data on attrition have not been collected yet and will be available at post-test. The level of attrition is low at this point. It remains to be seen if attrition could be a cause for concern with respect to the validity of the results.

Primary outcome

The analyses for the primary outcome will aim to answer the primary research question and will test whether the children who participated in the 1st Class@Number intervention differ at post-test significantly from those who did not participate in this intervention in the Quantitative Reasoning Test.

The variables used to answer the primary research question will be:

1. **Outcome variables**:
   Quantitative Reasoning score at post-test, due to be collected at the end of Year 2, in June 2017;

2. **Predictors:**
   (a) Pre-test: Quantitative Reasoning score at the end of Year 1 (June 2016);
   (b) School to which the child belongs;
   (c) Group allocation (intervention or control school).

Analysis

Primary analyses

All analyses will be conducted using 2-tailed significance tests at the 5% significance level.

Correction for clustering will be used via a ML model and the analyses will be conducted in SPSS/MLwiN or R (EEFAnalytics R package) using 2-tailed significance tests at the 5% significance level. Hierarchical linear models will be used to compare intervention and control groups on the post-test scores, controlling for pre-test scores; a two level model (pupils within schools) will be used to account for clustering at the school level. A random effects model will be used for school level.

Stratification for the randomisation will not be taken into account in the primary analyses, but the possibility that different training groups of TAs varied in their training efficacy will be explored by using a dummy variable for the identification of the six different training groups as a secondary analysis. Results from this analysis will be presented regardless of the effect. If this effect is significant, it will be taken into account in the interpretation and discussion of generalisability the results.

The design will be intent to treat and analyses will include the maximum number of participants. Currently we know that one intervention school interrupted the implementation of the intervention due to the trained TA's long-term illness, but the school has agreed to participate in the post-test, and the results will be included in the analysis. Reasons for other missing data will be investigated; if a high number is observed, possible biases will be investigated by analysing pre-test scores.

Analysis 1 will take into account the pre-test scores (fixed effect) and group allocation (fixed effect) as well as nesting in schools (random effect). Describing the equation with variable names:

$$\textit{Outcome measure}_{ij} = \beta_0 + \beta_1 \text{Pre-test}_{ij} + u_0 \text{School}_{ij} + \beta_2 \text{Group allocation}_{ij} + e_{ij}$$

β_0 intercept
β_1 slope for pre-test effect (as a fixed effect, this coefficient is not assumed to vary between schools)
u_{0j} slope for school effect (random effect) (also called level 2 residual by some)
β_2 slope of group allocation (Control vs Intervention at school level, fixed effect)
e_{ij} = level 1 residual

Effect sizes will be calculated using ANCOVA controlling for pre-test scores, to increase precision and power. Hedge's g will be used to calculate the effect size and will employ the means corrected for pre-test differences and the pooled standard deviation for both cells in

the design; the confidence interval will be reported using the traditional 95% interval. The intra-cluster correlation will be reported for pre- and post-test.

Missing data

Missing data will be scrutinised at the time of post-test, as only then will it be possible to be completely clear on its proportion with regard to pre-test and its nature. Presently, at the cluster level the proportion of missing data is small, so missing data does not seem to give cause for concern. Given the small number of pupils per school, every effort will be made to return to schools to collect data if pupils are absent on the testing day.

Besides analysing the proportion of missing data, we will consider its nature (missing completely at random, missing at random, or missing not at random) by investigating the reasons for missing data during the process evaluation. As no single analysis for missing data is definitive, every effort is being made during the trial to avoid attrition. The intervention and evaluation team are collaborating in this process. The offer by the intervention team of an alternative intervention is one of the recognised mechanisms for avoiding attrition. However, should there be a significant proportion of missing data (i.e. above 5%), or a pattern to it, measures will be taken to deal with attrition in the analysis. A logistic regression analysis to predict membership in the drop-out cell will be carried out using different variables (pre-test results; school descriptors such as above the median percentage of pupils eligible for school meals; intervention descriptors - trainer responsible for the school). If imputation is necessary due to the proportion of missing data, the relevant variables will be taken into account using multiple imputation techniques (e.g. pre-test data; eligibility for free school meals). Maximum likelihood estimation is a common method, but it relies on parametric assumptions and will be used if these assumptions are not violated. We will compare the results of the analysis using imputation and using all the data available, as imputation is seen as a remedy but not a cure for missing data.

Non-compliance with intervention

It is important to obtain information on non-compliance in order to avoid Type III errors. There were no cases of failure to take up the intervention. Compliance with the intervention in this study involves the quality of the training offered to TAs, the participation of TAs in the training, and children's attendance in the teaching sessions and the implementation of the intervention under the expected conditions.

- Description of variable(s) used to describe extent of intervention 'dosage' received

Fidelity of the training: Training delivered to TAs is highly scripted. All trainers were observed once and their delivery of the training was quite close to the script. The observations were scored but there is not sufficient variability for an analysis to take into account this variable. The outcome of this measure will be described as it supports the possibility of scaling up this complex intervention.

TA attendance variable: TAs' attendance to the training sessions was monitored; there was almost 100% attendance. The intervention team offered an extra session to a TA who had to miss two sessions. In one school the TA had to be replaced during the intervention period and the new TA attended the subsequent sessions. Thus there does not appear to be a need for compliance analysis at the TA level as there is no variability.

Pupil attendance variable: Data collected by TAs regarding pupils' attendance will define the dosage of the intervention received by pupils. Every effort is being made presently to obtain all the records. When information about pupils' attendance is obtained, it will be possible to

design a metric for the dosage received by pupils. Quartiles of absence per pupil can be used if the number of missed sessions varies sufficiently for a division into four groups, but a simpler division by the median will be used if number of missed sessions shows less variation. Although it is possible to use the number of missed sessions, this is likely to be a less sensitive measure because each teaching session starts with a brief review of the previous session. The pupil attendance variable will be used in a regression model to predict the outcome variable within the intervention group.

TA questionnaires for intervention schools: Further process evaluation data will be used to evaluate delivery of the intervention according to the theory of change developed by the intervention team. A questionnaire was developed for the TAs to indicate whether the delivery was according to protocol. It contains 12 items which ask the TAs to indicate fidelity (e.g. all lessons were delivered; there were three lessons each week; the TAs time for delivery was protected; a suitable area was provided; each of the mathematics topics was taught). The score in the questionnaire will be included in a structural equation model to assess whether it explains significant amounts of variance in the outcomes for the intervention group, as this would indicate that fidelity mediates the effect.

Additional analyses

TA questionnaire for control schools: TAs in control schools were asked to complete a questionnaire to help describe what "business as usual" means in this trial. The answers will produce dichotomous variables (e.g. was any additional support offered to the nominated pupils?) and continuous variables (e.g. number of topics which form the 1stClass@Number intervention that were taught to the nominated children either in class or in specific interventions). Our analyses will test whether outcomes in the control schools are associated with these variables (e.g. testing for a difference between schools that did put in place an alternative intervention for the nominated pupils and those that did not; testing for a difference between schools defined by the number of topics taught in 1stClass@Number that were taught in the control schools).

Subgroup analyses

- Subgroup analyses specified in the protocol

Two subgroup analyses were planned in the protocol: one in which the subgroups were defined in terms of eligibility for free school meals ((β_3 Eligible for free school meals (EVER6)$_{ij}$) and the other in which the subgroups were defined by gender. The focus on eligibility for free school meals is in line with the aims of the EEF. The analyses related to gender address a question frequently asked by teachers regarding effectiveness of maths interventions for girls; failure to address this question could result in less interest among schools in the implementation of the intervention in the future. In both cases, a main effect of sub-group membership and an interaction term will be added to the equations presented earlier on.

A multilevel model will be used, which will take into account the intra-cluster correlation. The formula is presented here with variable names for intelligibility.

For the subgroups defined by school membership:
$Outcome\ measure_{ij} = \beta_0 + \beta_1 \text{Pre-test}_{ij} + u_{0j}\text{School}_{ij} + \beta_2 \text{Group allocation}_{ij} + \beta_3\ \text{FSM}_{ij} + \beta_4 \text{Group allocation} * \text{FSM}_{ij} + e_{ij}$

For the subgroups defined by gender:

$Outcome\ measure_{ij} = \beta_{0j} + \beta_1 \text{Pre-test}_i + u_0 \text{School}_j + \beta_2 \text{Group allocation}_{ij} + \beta_3 \text{Gender}_{ij} + \beta_4 \text{Group allocation} * \text{Gender}_{ij} + e_{ij}$

β_0 intercept
β_1 slope for pre-test effect (as a fixed effect, this coefficient is not assumed to vary between schools)
u_{0j} slope for school effect (random effect) (also called level 2 residual by some)
β_2 slope of group allocation (Control vs Intervention at school level, fixed effect)
β_3 slope of free school meals in the top equation and of gender in the bottom equation
β_4 slope of interaction between group allocation and the previous term (free school meal or gender)
e_{ij} = level 1 residual

Effect size calculation

Hedge's g will be used in this project, as described previously.

Report tables

The tables used in the report will follow the EEF template.

References

Nunes, T., Bryant, P., Barros, R., & Sylva, K. (2011). The relative importance of two different mathematical abilities to mathematical achievement. *British Journal of Educational Psychology, 82*, 136–156.