Education Endowment Foundation

Catch Up® Numeracy

Evaluation report and executive summary

February 2019

Independent evaluators:

Jeremy Hodgen, Michael Adkins, Shaaron Ainsworth and Sheila Evans

UCL Institute of Education

The University of Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:
Jonathan Kay
Research and Publications Manager

Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP
p: 020 7802 1653
e: jonathan.kay@eefoundation.org.uk
w: **www.educationendowmentfoundation.org.uk**

About the evaluator

The project was independently evaluated by a team from the UCL Institute of Education and the University of Nottingham: Jeremy Hodgen, Michael Adkins, Shaaron Ainsworth and Sheila Evans.

The lead evaluator was Jeremy Hodgen.

Contact details:
Professor Jeremy Hodgen
UCL Institute of Education
University College London
20 Bedford Way
London WC1H 0AL
Tel: +44 (0) 207 612 6464
Email: jeremy.hodgen@ucl.ac.uk

Contents

Executive summary

The project

The Catch Up® Numeracy programme trains teaching assistants (TAs) to provide numeracy support to pupils in Years 3, 4, and 5 (ages 7–10). It aims to improve the numeracy of low attaining pupils. The support consists of two 15-minute one-to-one sessions led by TAs each week over a school year. In this trial, pupils were selected by schools based on maths attainment and the professional judgement of teachers. If participating pupils were judged to have reached an age-appropriate level of numeracy on the basis of termly assessment, they were 'rolled-off' the programme and no longer received support. Each school was asked to appoint a member of staff as the Catch Up® Numeracy Coordinator to coordinate the intervention and to support the TAs. Prior to the start of the intervention, TAs received three half-days of training and Catch Up® Numeracy Coordinators received one half-day of training. Catch Up® Numeracy was developed by the Catch Up® charity, part of the Caxton Trust, together with Dr Ann Dowker of Oxford University.

This was a randomised controlled trial involving 150 schools. Schools were randomly allocated to either deliver Catch Up® Numeracy or be in an 'active control' comparison group delivering a broadly time-equivalent amount of TA-led numeracy support sessions informed by the EEF's 'Making best use of teaching assistants' guidance report. The implementation and process evaluation involved: surveys of TAs, school coordinators, and senior leaders; analysis of delivery data; and case studies of intervention and comparison schools. The intervention began in September 2016 and finished in July 2017. Impact was measured using the Progress Test in Maths by GL Assessment.

Key conclusions	
1.	There is no evidence that Catch Up® Numeracy had an impact on pupils' maths outcomes, on average. This result has a low to moderate security rating.
2.	Among pupils eligible for Free School Meals, those who received Catch Up® Numeracy made two months less progress in maths compared to those who did not. This result may have lower security than the overall finding because of the smaller number of pupils.
3.	There was some weak evidence to indicate that children who received Catch Up® Numeracy developed a more positive attitude towards maths compared to those in the active control group.
4.	Schools implemented the intervention with a high level of fidelity. However, there was some mis-selection of pupils, which suggests that intervention programmes for low attaining pupils may not reach all pupils for whom they are intended.
5.	Teaching assistants reported that they valued the Catch Up® Numeracy training. However, there was no evidence that Catch Up® Numeracy changed the attitudes of teaching assistants to maths and maths teaching.

EEF security rating

Findings from this study have low to moderate security. The trial was an effectiveness trial, which tested whether the intervention worked under everyday conditions in a large number of schools. It was well-designed and well-powered. However, the following factors reduced the security of the trial: 17% of the pupils who started the trial were not included in the final analysis (for example, because they moved school between the intervention and the post-test); there were also some important differences in prior attainment between the pupils in Catch Up® Numeracy schools and those in the active control group; in addition, there was evidence of a 'floor effect' in the tests, which made it difficult to capture and compare achievement for lower attaining pupils (who are the target pupils for this programme), and thus harder to accurately estimate the size of the impact of Catch Up® Numeracy.

Additional findings

There is no evidence that Catch Up® Numeracy had a greater impact on pupils' maths outcomes than an equivalent amount of TA support, which is in line with the findings of the previous EEF trial (Rutt, Easton and Stacey, 2014). There was no evidence to suggest a negative impact on the attitudes or motivation of pupils and, indeed, the intervention may have had a small positive impact on pupil attitudes to maths. The implementation and process evaluation indicates that schools receiving Catch-Up® delivered it with high fidelity. The survey results also indicate that both Catch Up® Numeracy and the active control support were attractive to schools, and staff in both groups reported perceived improvements in pupil attainment and attitudes. In terms of the specific strengths of Catch Up® Numeracy, schools valued the structure of the programme and the opportunity for personalised learning for pupils. More than half of the schools in the active control group deemed it appropriate to deliver most or half of the sessions to groups of children, rather than individual pupils, while the Catch Up® Numeracy sessions were almost entirely individual, as recommended. Analyses across the Catch Up® Numeracy and the active control support indicated that group delivery did not appear to disadvantage pupils.

Cost

The average cost of Catch Up® Numeracy is £39 per pupil per year over three years. In addition, nine half-days of supply cover were required for attendance at training (three for each TA and three for the school coordinator).

Table 1: Summary of impact on primary outcome

Outcome/ Group	Effect size (95% credible interval)	Estimated months' progress	EEF security rating	No. of pupils	EEF cost rating
Progress Test in Maths	-0.04 (-0.21, 0.13)	0	🔒🔒🔓🔓🔓	1481	£ £ £ £ £
Progress Test in Maths FSM pupils	-0.14 (-0.33, 0.09)	-2	N/A	551	£ £ £ £ £

Introduction

Background evidence

Over the past 20 years, the use of teaching assistants (TAs) has dramatically increased in England (Blatchford, Bassett, Brown and Webster, 2009; Department for Education, 2017), and schools now spend more than £4.4 billion annually on TAs (Sharples, Webster and Blatchford, 2015). However, whilst TAs now play a very major role in education, ways of deploying and using TAs to support learning effectively are poorly understood. TAs are frequently used to support low attaining pupils (Warhurst, Nickson, Commander and Gilbert, 2013). There is an increasing body of evidence indicating that the receipt of TA support often has no, or negative, effects on pupil attainment (Blatchford et al., 2009). However, there is some evidence that some well-designed interventions delivered by TAs may have an effect on student attainment and several EEF-funded trials have shown some promise, although these are mainly in the area of literacy (Sharples, 2016; Sharples et al., 2015). More recently, Pellegrini, Lake, Inns, and Slavin's (2018) meta-analysis of effective tutoring programmes in mathematics demonstrated positive benefits for targeted tutoring interventions delivered by TAs, and, indeed, found that tutoring by TAs was as beneficial as tutoring by qualified teachers.

Catch Up® Numeracy, the focus of this evaluation report, is one such intervention delivered by TAs that has been judged to have promise as a result of a previous trial (Rutt, Easton and Stacey, 2014). Catch Up® Numeracy is a research-based intervention for pupils aged 6–14 who are struggling with numeracy (Holmes and Dowker, 2013).

Evidence of the efficacy of Catch Up® Numeracy comes from an earlier independent evaluation funded by the Education Endowment Foundation and conducted by the National Foundation for Educational Research (Rutt, Easton and Stacey, 2014) and several studies conducted by the developers. This existing evidence relates to a 'fixed-time' implementation of Catch Up® Numeracy without a progress review and the opportunity for 'roll off' for individual pupils, as in the current trial.

In the previous study by Rutt et al. (2014), 336 pupils in Years 2 to 6 from 54 schools participated in a three-arm randomised controlled trial, each randomly assigned within their school to one of three groups: the Catch Up® Numeracy intervention, a 'matched-time' group in which they received the same amount of one-to-one maths instruction with a TA but not using Catch Up® Numeracy, and a 'no-intervention' group where pupils received no additional TA support beyond normal classroom instruction (Rutt et al., 2014). All three groups sat the Basic Numeracy Screening Test (Gillham and Hesse, 2001) before and after the intervention. After attrition, the analysis involved 108, 102 and 108 pupils for each of the Catch Up®, matched time, and no-intervention groups, respectively. Using an intention-to-treat analysis, the independent evaluators found that Catch Up® and the matched-time group made greater progress when compared to the business-as-usual control: effect sizes (Hedges g) of +0.21 (CI 0.42–0.01) for Catch Up® and +0.27 (CI 0.49–0.06) for the matched-time group. These differences were statistically significant, but there was no significant difference between the Catch Up® and the matched-time groups. However, there may have been some cross-contamination between these groups, because delivery was by TAs within the same schools.

An earlier study conducted by the developers found positive effects for the Catch Up® Numeracy intervention. Dowker and Holmes (2013) conducted a quasi-experimental study involving 440 pupils from 15 local authorities in England and between two and six schools in each local authority. Each school identified two to four pupils who were considered to have difficulty with arithmetic. Pupils were randomly assigned by their teacher to either Catch Up® intervention (N = 348), matched-time one-to-one TA support (N = 50), or no-intervention (N = 42) groups. The researchers found a statistically significant difference in gain on the Basic Number Screening Test (Gillham and Hesse, 2001) between Catch Up® and both the matched-time and no-intervention groups with effect sizes (Cohen's *d*) of +0.47

and +0.55, respectively. In an earlier report of the same study (Dowker and Sigley, 2010),[1] the effects were compared using mean ratio gains. The mean ratio gains were 2.2, 1.47 and 1.25 for the Catch Up®, matched-time, and no-intervention groups.

Although these studies provide some evidence of the effectiveness of Catch Up® Numeracy, there is a lack of firm evidence as to whether it is more effective than normal 'good practice' of numeracy support by TAs. Hence, there was sufficient equipoise, or justification, to conduct a further randomised controlled trial to estimate the impact in comparison to an active control group.

Intervention

Catch Up® Numeracy is a one-to-one intervention delivered by teaching assistants for learners who are struggling with numeracy. The intervention is described below. Further detail is available in the Catch Up® Numeracy file (Catch Up®, 2017).

The intervention evaluated in this trial differs from the version evaluated in the previous trial (Rutt et al., 2014). In the trial evaluated here, pupils were assessed on a termly basis to identify their areas of weakness and their overall eligibility for the intervention. Pupils who were judged to have reached an age-appropriate level of numeracy were then 'rolled off' the intervention and no longer received the intervention. This is how Catch Up® Numeracy is usually delivered.

There are two issues to highlight about the intervention that have implications for the evaluation and are discussed later in this report:

- The use of delayed post-tests to evaluate impact: since the intervention aims to enable pupils to 'catch up' by delivering sustained increases in attainment in numeracy, the intervention was evaluated using a delayed post-test administered in the autumn term following the academic year in which the intervention was implemented. The delayed post-test had the advantage of capturing sustained gains in attainment. However, this meant that testing had to take place after the summer break. Although there is a widespread view that pupils' attainment falls back over the summer, Brown et al.'s (2008) longitudinal large-scale survey of progression in primary indicates that this is not the case; across the primary age range, on average, pupils made gains over the summer period and these gains were proportional to the gains made over the remaining school year. However, this delay might lead to higher attrition because some pupils may have moved schools between school years.
- The guidance provided to schools on pupil selection: normally, guidance and support is provided to schools on the selection of pupils during Catch Up® Numeracy training. In this trial, the selection of pupils was carried out prior to randomisation and before training in order that selection was not influenced by group allocation. As a result, the schools in this trial had more limited guidance on how to select pupils and did not follow the usual Catch Up® Numeracy practice of using a standardised test as part of the selection process. Misallocation to groups by attainment is a general problem for schools (see, for example, Francis, Archer, Hodgen, Pepper, Taylor and Travers, 2017; MacIntyre and Ireson, 2002). Thus, this limited guidance and support might lead to some middle and high attainers being selected for whom the Catch Up® Numeracy intervention may not be suitable, and might also contribute to imbalance within the sample.

[1] The ratio gain is calculated by the months gained in Mathematics Age divided by duration of intervention in months. The mean ratio gains reported here approximate to Cohen's *d* of around 0.3 and 0.4 for Catch Up® compared to the matched-time and non-intervention groups.

The short description below follows the Template for Intervention Description and Replication (TIDieR) checklist,[2] which should be read in conjunction with the logic model (see Figure 1):

1. Brief name: Catch Up® Numeracy.

2. Why (rationale/theory): Catch Up® Numeracy is a research-based intervention targeted at children who are low attaining in mathematics (Holmes and Dowker, 2013). The intervention is guided by a componential approach to numeracy (Dowker, 2009) and is designed to address individual children's particular difficulties and weaknesses through one-to-one support sessions delivered by teaching assistants (TAs). This trial follows an earlier Education Endowment Foundation study that showed some positive effects for the Catch Up® Numeracy intervention (Rutt et al., 2014). The intervention breaks numeracy down into ten components and at the start of the intervention assesses children's ability on each. By using a checklist approach based on this assessment, the teaching assistant's subsequent instruction is targeted to always address the child's exact area of weakness (Dowker and Sigley, 2010). Components include counting verbally, counting objects, derived facts and so on. The approach is based on research indicating that numeracy is not a single 'big' skill, but a compound of several 'little' skills that seem to be quite discrete (Dowker, 2009). Children (and adults) may be very strong in some skills but very weak in others. By recognising and building on this finding, the Catch Up® Numeracy intervention enables tutors to diagnose and treat problems precisely and effectively. It is anticipated that the intervention will impact on children's attitudes towards mathematics as well as their attainment in numeracy.

3. Who (recipients): For this trial, the target group was primary school pupils in Years 3, 4 and 5, whom the school had identified as struggling—those amongst the lowest attaining 15–20% of children for their age. Schools are based in the North East, Yorkshire, the Peterborough area, and London. The restriction to primary pupils was for this trial only. The general Catch Up® Numeracy intervention is aimed at learners aged 6–14. In addition, the process of selection and assessment was different for this trial, and, to ensure allocation was blind to treatment, the selection of pupils was made by schools before training and randomisation. For the usual Catch Up® Numeracy intervention, selection is made after training, and schools receive guidance and training on the selection of pupils (Catch Up®, 2017).

4. What (materials): TAs were supplied with material detailing how each session should be structured, and suggestions on its delivery. Additional activities including lesson plans were available via the Catch Up® online platform.

5. What (procedures): Catch Up® Numeracy is a one-to-one intervention for learners who are struggling with numeracy that aims to improve attainment in numeracy and attitudes to mathematics. It consists of two 15-minute sessions per week that are delivered by TAs. At the beginning of the intervention, a structured assessment is conducted with each pupil in order to identify the pupil's componential weaknesses and appropriate number range for the intervention. The assessment is normally expected to take between 60 and 90 minutes, although this will vary according to each pupil's age and prior knowledge. Catch Up® advises completing the assessment phase as quickly as possible without overloading the pupil. The assessments should be carried out one-to-one by the TA who will be working with the pupil since this is part of the engagement process. To prepare them for delivering the intervention, TAs received three half-day training sessions, each involving a follow-up task in school. An additional optional 'interim' training session was provided during the implementation year for TAs (and school coordinators). Accreditation for TAs is available through the optional 'Delivering Catch Up® Numeracy' qualification. Unlike in the previous independent trial (Rutt et al., 2014), the intervention was delivered flexibly and pupils 'rolled off' the intervention (following the usual delivery model for Catch Up® Numeracy). All pupils in the intervention were assessed on a termly basis as to their eligibility for the intervention. Pupils who were judged to have reached an age-appropriate level of numeracy then no

[2] http://www.bmj.com/content/348/bmj.g1687

longer received the intervention. Extensive guidance is provided for schools and TAs in the Catch Up® Numeracy file (Catch Up®, 2017).

6. Who (implementers): Each school identified two TAs who each delivered Catch Up® Numeracy support to six pupils. Each school also appointed a member of staff as the school's Catch Up® Numeracy Coordinator to coordinate the intervention in school and to support the TAs. The coordinator was normally a teacher and was often the school SENCO or a member of the school's leadership team. S/he was responsible for monitoring and reviewing the delivery of the intervention, ensuring delivery records are up to date, and providing ongoing support for the TAs. To prepare them for their role, coordinators received training alongside the TAs.

7. How (mode of delivery): Two 15-minute one-to-one sessions per week.

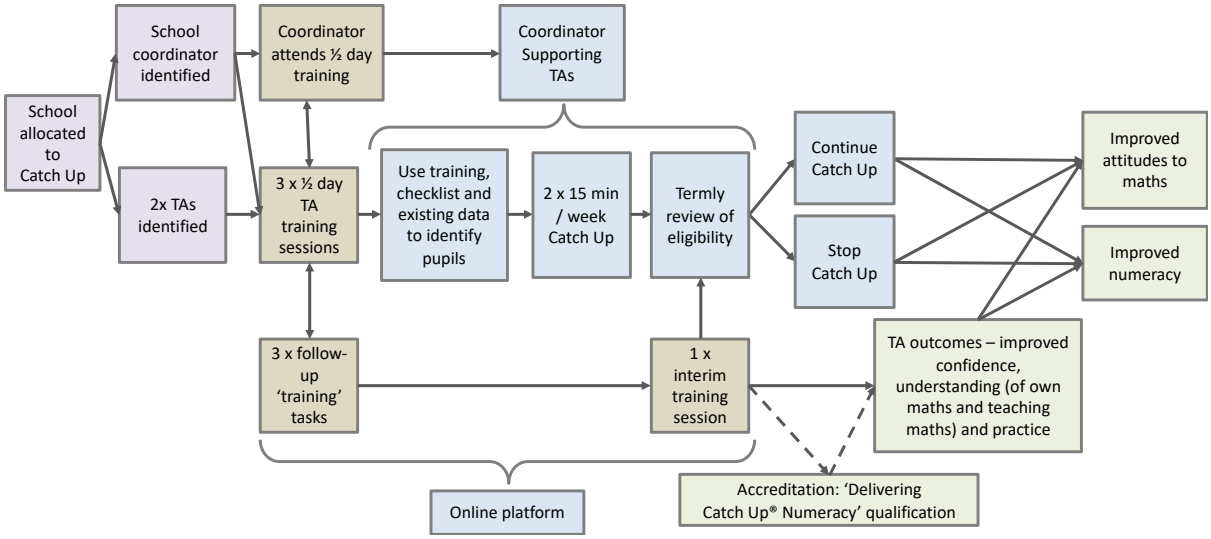
8. Where (setting): The one-to-one sessions are delivered in a location outside the classroom.

9. When and how much (dosage): Children received one, two, or three terms of support depending on if, and when, they rolled off the intervention.

10. Tailoring: TAs are allowed to substitute their own activities provided the substitute activities are directed at the same componential weakness as the intervention materials and were integrated into the standard structure of a session.

11. How well (planned): Effective implementation requires TAs to be trained before they deliver the intervention, and school coordinators to coordinate and support intervention delivery (including ensuring that pupil records are maintained).

Figure 1: The logic model for Catch Up® Numeracy



Active control

Active control schools were asked to appoint two teaching assistants to provide regular one-to-one numeracy support sessions (totalling 30 minutes per week) to pupils struggling with numeracy. Pupils, TAs, and school coordinators were identified prior to randomisation. Schools were supplied with EEF guidance on making best use of TAs (Sharples et al., 2015) and were advised to organise an information session on this guidance. All pupils should have the one-to-one sessions at the start of the school year. Schools were asked to review progress each term, rolling off approximately one third of the participating pupils at each time point—end of autumn, spring and summer terms.

Evaluation objectives

The aim of the project was to test and evaluate, within a rigorous and high quality research framework, the impact on underperforming Year 4 and Year 5 pupils of the Catch Up® Numeracy intervention on attainment and attitudes to mathematics, when delivered for a period of up to three terms by trained TAs against 'in-house approaches' to supporting similar pupils informed by best practice guidance (Sharples et al., 2015) when delivered by TAs for a similar time. Although the intervention is targeted at a wider age range of 6–14 years, it was judged that equating attainment across a narrower age range would produce a more robust outcome measure. Primary was chosen because the intervention is more widely used in primary, and Year 4 and Year 5 were chosen to avoid a clash with preparation for National Assessments in Year 6. More detailed research questions (RQs) were outlined in the evaluation protocol, which is published on the EEF website alongside a detailed statistical analysis plan (SAP).[3]

The primary research question was:

1. Does Catch Up® Numeracy have a significant effect on pupil attainment in mathematics when compared to an active control?

The impact evaluation addressed the following secondary research questions:

2. Does Catch Up® Numeracy have a significant effect on children's attitudes towards mathematics when compared to an active control group of schools delivering matched time TA support?
3. Are the effects on attainment and attitudes different for children eligible for free school meals?
4. Are the effects on attainment and attitudes different for girls and boys?
5. To what extent are any effects on attainment and attitudes mediated by the treatment time?

In addition to the standard investigation of implementation dimensions and factors (Humphrey, Ledrum, Ashworth, Frearson, Buck and Kerr, 2016), the implementation and process evaluation addressed the following research questions specific to the intervention:

6. To what extent do schools, coordinators, and teaching assistants perceive the Catch Up® Numeracy professional development for teaching assistants and coordinators to be effective?
7. To what extent do the Catch Up® Numeracy intervention schools, teaching assistants, and coordinators adhere to the guidance and materials?
    i. How variable is the quality of implementation in the intervention schools?
    ii. To what extent are the Catch Up® Numeracy resources sufficient, appropriate, and easy to access, and is the assessment guidance sufficiently flexible and child-friendly?
    iii. What school and contextual factors afford or constrain the quality of implementation?
    iv. In what ways do schools manage and support the teaching assistants?
    v. To what extent, and how, do schools enable dialogue to take place between the teaching assistants and the relevant class teachers?

[3] Both the evaluation protocol and the SAP are available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/catch-up-numeracy-2015/

8. Are children assessed for eligibility on a termly basis in both intervention and control schools and do children judged to have reached an age appropriate level of numeracy 'roll off' the intervention?
9. What does usual practice in control schools look like?
    i. How, and to what extent, do the active control schools implement the Education Endowment Foundation's 'Making best use of teaching assistants' guidance?
    ii. To what extent does the half-day planning session enable schools to implement the guidance?
10. To what extent does Catch Up® Numeracy have an effect on teaching assistants' attitudes towards mathematics and mathematics teaching when compared to an an active control group of schools delivering matched time TA support?

Ethics and trial registration

The study was reviewed and approved by the University of Nottingham School of Education (ref: 2016/989/CD; date granted: 10 March 2016). Approval was granted on the basis of opt-out consent by pupils and their parents/carers. Pupils and their parents or carers were provided with information about the project and how personal data would be processed and given the opportunity to object and either withdraw their data from the trial or to withdraw from the programme of support entirely, of the research and testing element, or from the programme entirely. Surveys (for TAs, coordinators, and senior leaders) included information on the research and these staff were informed that completion of the survey would be taken as giving consent. For case studies, observations and interviews with TAs, and other school staff, active consent was sought on an opt-in basis.

The headteacher of each school was required to sign a Memorandum of Understanding prior to randomisation in order to participate in the trial (see Appendix C).

The trial is registered with ISRCTN (www.controlled-trials.com) reference number: ISRCTN15428227.

Data protection

Participants, and pupils' parents or carers, were notified in the information sheet giving them the right to object. Specifically, these letters and information sheets highlighted that they could choose not to take part and stop at any time; that they could ask for data to be removed from the study up to one month after the end of the data collection; and that data would be linked to the National Pupil Database, other official records and shared with the developers, the Department for Education, Education Endowment Foundation, the data contractor Fisher Family Trust and in an anonymised form to the UK data archive.

The data were processed under Schedule 2 of the Data Protection Act 1998. In this, one of the conditions for fair processing of personal data needs to be fulfilled. In this case, Condition 1 has been met in that the data subject has given their consent to processing which applies here as parental opt-out consent was obtained, as consent does not have to be explicit. These arrangements were reviewed and approved as part of the ethical approvals process at the University of Nottingham. Note that the data collection and processing arrangements were carried out prior to the introduction of the GDPR.

Project team

Development and delivery team:

Dr Graham Sigley (Deputy Director, Catch Up®) directed all aspects of the project that came within the responsibility of Catch Up®/the Caxton Trust.

Dr Ann Dowker (University of Oxford) provided support, information, advice, and guidance in maximising the effective implementation of the project on a consultancy basis.

Julie Lawes (Director, Catch Up®) and later **Theresa Rogers** (Chief Executive Officer) provided strategic input and support where appropriate and deputised for Dr Graham Sigley where required.

Ann Fletcher (Administrator) assisted the development and delivery of the project, including maintaining contact with schools.

Catch Up® Accredited Trainers (CUATs) delivered all training and briefing sessions and contributed to the production of project briefing material that was required in addition to the training and support materials that were included within the Catch Up® Numeracy integrated training, resource, and support package.

The Catch Up® office team provided input and support in the delivery of the project, including the ongoing support that project schools and their staff are entitled to as part of their membership of the Catch Up® Community resulting from undertaking Catch Up® Numeracy training.

The evaluation team:

Professor Jeremy Hodgen (University of Nottingham, until 31 August 2017; UCL Institute of Education, from 1 September 2017): Principal investigator.

Professor Shaaron Ainsworth (University of Nottingham): Implementation and process evaluation lead, advice on design and statistical methods.

Dr Michael Adkins (University of Nottingham): statistician, impact evaluation, assistance with day-to-day project management.

Dr Sheila Evans (University of Nottingham): Implementation and process evaluation, case studies, contact with schools.

Kanchana Minson (University of Nottingham): Project administrator, contact with schools, GL Assessment and the developer.

Tamsyn Smith (University of Nottingham): Administrative assistant.

Methods

Trial design

The trial utilised a two-arm randomised controlled trial comparing the Catch Up® Numeracy intervention to an 'active control' with randomisation at school level and a delayed post-test of attainment in mathematics. The aim was to recruit 1,800 low attaining pupils from 150 schools (12 per school), to be supported by two TAs in each school. Pupils and TAs were identified prior to randomisation. A pre-test was administered to all pupils prior to randomisation (see **Table 2** for a summary).

The previous Catch Up® Numeracy trial (Rutt et al., 2014) randomised at the individual level, with the six pupils per school being equally split into three arms ('business as usual' control, active 'matched-time' control, and intervention). However, this design was potentially susceptible to cross-contamination between the matched-time and Catch Up® arms. As a result, this trial proceeded on the basis of school-level randomisation to avoid any potential of issues of contamination.

Table 2: Summary of trial design parameters

Trial type and number of arms		Effectiveness trial with two arms: Catch Up Numeracy and an active control
Unit of randomisation		School
Stratification variable(s) (if applicable)		Region: North East, Yorkshire, Peterborough, and London
Primary outcome	variable	Delayed test of mathematics attainment
	measure (instrument, scale)	Age-standardised mathematics score (GL PTM age appropriate test, scale based on standardised scores)
Secondary outcome(s)	variable(s)	Attitudes to mathematics
	measure(s) (instrument, scale)	Bespoke instrument

As already noted, the previous three-arm trial had already indicated that both Catch Up® and the active control condition could improve attainment of pupils by a roughly equivalent effect size (0.21 and 0.27, respectively) against a business as usual control condition. Hence, following discussion amongst the developer, funder, and evaluator, it was felt that there was a need to focus on evaluating Catch Up® Numeracy against an active control matched time equivalent in a large trial setting rather than developing an optimised but repeated design of an existing trial. A three-arm trial was considered but was judged to present potential recruitment difficulties.

In the active control arm, schools were asked to deliver equivalent time one-to-one numeracy support by TAs (30 minutes per week) and were provided with the EEF's 'Making Best Use of Teaching Assistants' guidance report (Sharples et al., 2015). As with the Catch Up® Numeracy intervention, all pupils were to receive support from the start of the academic year.

Schools in both arms of the trial were asked to carry out individual termly reviews, rolling off any pupils judged to have reached an age-appropriate level of numeracy on the basis of termly assessment, and thus would no longer receive support. On the basis of the developer's previous experience, schools

were advised that approximately one third of the participating pupils would roll off at each time point—end of autumn, spring and summer term—in both intervention and control arms.

Various post-test timing options were considered, including immediate post-tests at roll-off, end of year tests, and delayed post-tests. None of these options was ideal. Whilst the use of multiple testing points was considered to be the most robust in terms of understanding the effect on pupil attainment, multiple testing was judged to be too burdensome for schools and pupils. Following discussions with the developer and the EEF, it was decided that a delayed post-test in the term following the end of the intervention was the best compromise in that all pupils would be tested at the same time point and there would be a delay for all pupils, thus enabling us to capture the sustained effects of the intervention that were judged to be particularly important for the target group of low attaining pupils.

All schools were provided with an incentive payment of £450 once they had completed the post-tests. Intervention schools were provided with Catch Up® Numeracy training, but in order to avoid contamination of the sample for long-term analysis, all control schools were provided with Catch Up® Literacy training in the form of a waitlist in September 2017. They were requested not to use the training on children identified as eligible for Catch Up® Numeracy.

Participant selection

The developer was responsible for recruiting schools to the intervention. Catch Up® sent out mass mailings to primary schools in three regions: Yorkshire, North East, and Peterborough. Due to recruitment difficulties, a fourth region (Havering, London) was added later. Schools were asked to submit an expression of interest and Catch Up® followed this up with several recruitment events in which both the developer and evaluator provided potential applicants with information on the trial.

All state primary schools were eligible as long as they had not already purchased Catch Up® Numeracy and could provide a minimum of six Year 4 and six Year 5 pupils who were eligible for the intervention (or alternatively four eligible pupils in each of Years 3, 4 and 5). Junior schools could be included if they agreed to get the pre-test data from the eligible Year 3s (then current Year 2s) from their partner infant schools. Pupils were eligible as long as they were in the bottom 15–20% of mathematics achievement based on attainment at KS1, as well as the professional judgement of teachers based on their more recent performance.

In addition, in order to be entered into the randomisation the schools had to provide:

- a signed Memorandum of Understanding;
- confirmation that consent forms had been sent out to parents/carers and any opt-outs;
- pupil data for those identified as eligible (UPN, forename, surname, date of birth, sex, free school meal eligibility, and teaching assistant assignment); and
- pre-test data for all eligible pupils.

Outcome measures

The primary outcome measure was pupil test scores using the Progress Test in Mathematics published by GL Assessment, which assesses mathematics more generally and includes items related to shape, data handling, and algebra in addition to the number and numeracy of the intervention. For pre-testing, the pupils took the age-appropriate paper and pencil versions of the test: PTM 7, 8 and 9, respectively. Depending on recruitment, pre-tests were administered either in June or July 2016 when pupils were in Year 2, 3 or 4, or in September 2016 when pupils were in Year 3, 4 or 5. For the delayed post-test (administered in October, November and December 2017), the pupils, who were then in Year 4, 5 and 6, took the age-appropriate paper and pencil versions of the test: PTM 8, 9 and 10, respectively. The

pre-test was administered by schools before randomisation, with delivery and marking by GL Assessment. For the post-test, these were administered by independent invigilators employed by the evaluation team and were then marked by GL Assessment. The invigilators and GL Assessment were blind to the intervention and randomisation allocation.

GL Assessment was consulted and advised that the relevant PTM tests were appropriate for pupils at all attainment levels including the low attaining students targeted by Catch Up® Numeracy.[4] However, after the study design was agreed and the first version of the evaluation protocol had been published, some unpublished re-analysis of the results of previous trials suggested that the use of age-standardised scores may in certain cases inflate effect sizes due to floor and ceiling effects. Hence, this was investigated by conducting an initial descriptive analysis of the histograms of the raw and age-standardised pre- and post-test scores to check whether floor and ceiling effects are evident, and whether there were differences between the raw and age-standardised scores. Substantial floor effects were found on the age-standardised scores. This is discussed later in the report under 'Distribution of the PTM tests'. In addition, the model was fitted to the raw scores, adding an additional non-nested random effect for test-level. A Bayesian equivalent of Tobit model (McBee, 2010), which incorporates adjustments for floor or ceiling effects, was also fitted using the standardised scores.

The secondary outcome measure was the pupil score on an attitudes to mathematics survey using an amalgam of four single-scale items that were combined by summing, developed from previous work by Dowker (Thomas and Dowker, 2000; Krinzinger et al., 2007). These were judged to be quick and cost-effective to administer. Although the measure used was not specifically validated, similar short attitude scales using similar items drawn from Dowker's work have been shown to be valid and reliable (for example, Núñez-Peña, Guilera and Suárez-Pellicioni, 2014). This measured self-rating and liking for mathematics. In addition, in order to assess the extent to which any effects are specific to mathematics rather than relating to more general attitudes to learning, pupil attitudes to reading were measured using parallel items.

Sample size

The trial was designed to maximise the possibility of detecting a small effect size, particularly given the results of the previous trial and the lack of a 'business as usual' control. Raudenbush et al.'s (2011) Optimal Design software was used to run a series of statistical power calculations on the basis of recruiting two TAs and 150 schools for a three-level cluster randomised trial with the intervention delivered at level 3 (that is, the school level). The structure of the intervention was made up of three levels—pupils were clustered in TAs who were further clustered in schools.

Multilevel models are particularly useful with cluster randomised designed trials as these models take account of the natural clustering that takes place in a school. Schools, in particular, form natural hierarchical structures providing their pupils with a very specific experience that differs from other schools in the area. This shared experience often translates into subtle differences in outcomes at an aggregated level. To be specific, pupils taught by a specific TA will likely have a more similar experience than other pupils taught by a different TA in the same school, or pupils taught at different schools. Taking this clustering into account allows us to appropriately adjust the standard errors of our estimates to avoid making stronger conclusions than are supported by the evidence.

The number of pupils per TA (two, four, and six) was varied in order to provide us with a realistic estimate of power given the expected pool of participants in schools and fixed the remaining parameters. These were as follows: $\alpha = 0.05$ (which refers to the probability of rejecting the hypothesis tested when it is true—5%), and intra-cluster correlation (ICC) for level 2 (TAs) = 0.05 and for level 3 (schools) = 0.10 (which refers to the variance at the TA and school levels respectively). An additional

[4] Technical information on the tests and the age standardisation process is available on the GL Assessment website: https://www.gl-assessment.co.uk/support/ptm-product-support/

pre-test covariate was also included as a school-level aggregate with the assumption that the post- and pre-test have a correlation of 0.80, setting the level 3 variance explained at $0.80^2 = 0.64$. This has the effect of reducing the overall variance and boosting the expected statistical power of the study.

As noted in the statistical analysis plan, there were few existing trials funded by the EEF that had three levels of variation and had used the PTM test, so the assumptions came with some uncertainty. The results of the calculations before randomisation can be seen in Figure 2 and Table 3. At 80% power, where under repeated testing the interval would be expected to contain the 'true' effect 80% of the time, the minimum detectable effect size (MDES) is 0.172 with six participants per TA. Comparing the minimum detectable effect sizes (measured in standard deviations), the calculations at randomisation underestimated the variation at the TA and school level, which was significantly higher than anticipated (see Table 7). This may be due to schools interpreting the eligibility criteria very broadly and the weaker than expected correlation between differing levels of the PTM test for these pupils. (See Table 7 in the analysis section for full details of the MDES at design, randomisation, and analysis stages.)

Figure 2: Three power curves highlighting the change in the minimum detectable effect size when the number of pupils per TA is varied (Black = 2, Red = 4, Blue = 6)

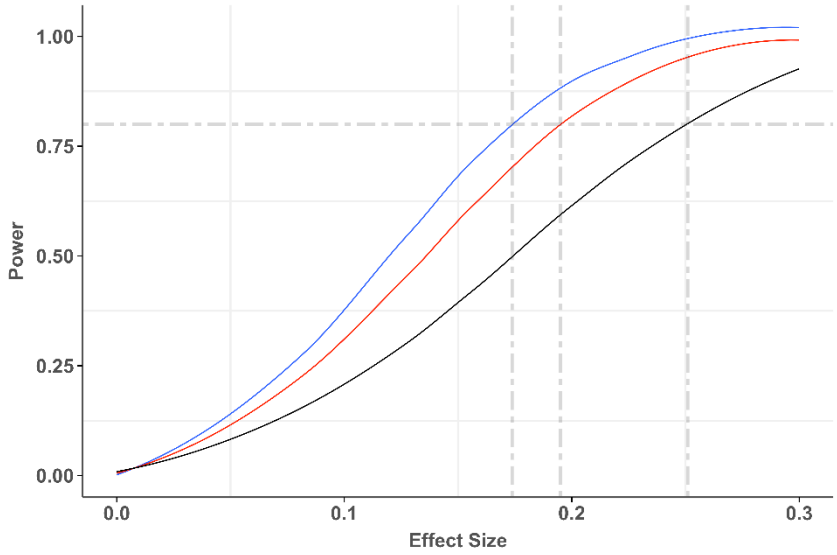


Table 3: Power analysis results at design stage for outcome only and with PTM covariate; estimates are subject to rounding

Catch Up® power analysis—all pupils (including FSM)			
	2 pupils per TA	**4 pupils per TA**	**6 pupils per TA**
Outcome only	0.27	0.22	0.20
PTM covariate	0.25	0.20	0.17

For FSM pupils, based on an assumption that approximately 35% of the nominated pupils would have been eligible for free school meals during their primary schooling (everFSM), this provided for approximately two children per TA and four children per school. Table 4 shows the MDES for FSM pupils based on the same assumptions. (See Table 7 for full details.)

Table 4: Power analysis results for FSM pupils only at design stage on both an outcome only and with PTM covariate; estimates are subject to rounding

Catch Up® power analysis—FSM pupils only	
	2 pupils per TA
Outcome only	0.27
PTM covariate	0.25

Randomisation

Randomisation and the statistical analysis were carried out by the statistician, Michael Adkins. In order to ensure that analysis was blind to group allocation, the dataset was blinded by another member of the evaluation team, Kanchana Minson. After data cleaning, but prior to the analysis, identifying features of the dataset were blinded. This included the treatment condition, pupils, TAs, and schools.

Randomisation took place in July and September 2016, with baseline testing taking place in June prior to schools being randomised. Due to recruitment difficulties, schools were randomised in three batches: 135 in May 2016, seven in June 2016, and nine in September 2016.

Randomisation used the following procedure. First, the school-level data was pre-processed—checking school names, Unique Reference Numbers, and postcodes against Edubase records. Second, a split function was set up in R which separately randomised the four regional blocks (R code included in the SAP)[5] and then worked on the basis of simple randomisation to split schools equally into intervention and active control arms. Third, simulations were run to check that any imbalances between the groups were likely to fall within acceptable levels. The level of imbalance was examined in publicly available school-level indicators 2015 data—percentage of everFSM, average KS2 point score, and percentage of pupils with English as an additional language. These simulations indicated that imbalances on average KS2 score and proportion of EAL students was approximately 2%, and for everFSM, less than 6%. Fourth, the schools were randomised using a randomly generated seed. Finally, as a double-check, a balance analysis was conducted to confirm that imbalances fell within the bounds predicted by our simulations.

Statistical analysis

The impact analysis investigated the effect of Catch Up® Numeracy against the active control condition on the basis of intention-to-treat using a four-level linear multilevel regression model of pupils nested in TAs and nested in schools, estimated by Bayesian Inference. This is a change to the original intention as reported in the SAP because accounting for randomisation group made a small, but practically significant, difference to the precision (or credible interval) around the effect.

Model fit was assessed by leave-one-out cross-validation (LOO-CV) and the model specification is presented and discussed in the analysis section. Intra-cluster correlation coefficients—designed to determine the amount of variance at each level—were estimated using variance components models that were empty of covariates and estimated for the post-test, post-test with missing data imputation, the pre-test, and also the KS1 average point score.

The secondary analysis examined pupil anxieties and attitudes to mathematics on the basis of intention to treat using a bespoke instrument designed to assess how much the pupils liked mathematics and

[5] The SAP is available here: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/catch-up-numeracy-2015/

how hard they found studying mathematics. This was modelled in the same way as the primary outcome, although a multivariate multilevel model that regressed the primary and secondary outcome measures simultaneously was also fitted as part of the sensitivity analysis.

Subgroup analyses

Modelling was also conducted to allow the comparison of the impact of Catch Up® Numeracy using the primary outcome measure on free school meals and non-free school meals using separate interaction terms against the treatment condition. In addition, Dowker, Sarkar and Looi's (2016) review suggested that boys may rate themselves as stronger in mathematics in comparison to girls' self-rating, and this might result in a differential impact on both attitude and attainment. Hence, modelling was also conducted to compare male and female pupils. Each of these analyses was fitted separately, and on the basis of intention to treat.

Further analyses

Further analyses of the primary outcome incorporated compliance data, as well as school-level variables such as the KS1 average and KS2 average to adjust for imbalance in the groups. The intention had been to add a group-level variable to assess the impact of the TAs maths level, but after data collection, it was found that there was little variation with the majority of TAs having achieved an A*–C GCSE maths qualification or an equivalent.

Effect size calculations

The effect sizes of the treatment were calculated using a different process to those set out in the SAP because developments in the STAN software allowed for a more efficient and less error-prone process. This process is functionally equivalent and produces the exact same results as the process in the SAP. Instead of using a generated quantities section within the modelling, the posterior samples were extracted and the difference between the treatment conditions was divided by the square root of total variance. Summary statistics were then calculated to derive mean, standard deviations, and 95% Bayesian credible intervals. In a Bayesian framework, there is no direct equivalent to null hypothesis testing. Following Kruschke and Liddell (2017), a ROPE (Region of Practical Equivalence) analysis was carried out in order to examine whether the null hypothesis should be accepted as credible or practically significant. This procedure examines the proportion of the Highest Density Interval (HDI) that falls within a pre-determined effect size. Again, following Kruschke and Liddell (2017), this effect size was set at ±0.1 around 0.

Sensitivity analyses

The primary and secondary analyses were subjected to several robustness checks. While a deviation from the SAP, on receipt of the pre- and post-tests, the distributions of the test scores were examined descriptively and found to have thicker tails than expected from those with a normal Gaussian distribution. In addition, there were significant floor effects for both sets of age-adjusted test scores in that—for a substantial proportion of pupils—their age-adjusted score was at the minimum (see Appendix E, Figure 1). These floor effects were much more significant than had been anticipated in the SAP. As a check on the robustness of the findings, two parameters were added to the model. The assumption that the tests were student-t distributed added the ν parameter, which estimated and adjusted for the departures from normality, and a censoring parameter identified those observations within the likelihood that were censored on the left of the distribution, adjusting as much as possible for the floor effects within the outcome. The latter adjustment is a Bayesian equivalent of the classical Tobit model.

The findings of the Bayesian modelling were checked for robustness by using alternative prior distributions, comparability with classical estimates, and software issues by fitting within lme4 in R, and

MLwiN in using two approaches—Iterative Generalised Least Squares (IGLS), which is a classical Maximum Likelihood based algorithm, and the Markov Chain Monte Carlo (MCMC) based Gibbs sampler. In MLwiN, the default priors are generally uninformative, with improper uniform priors being used for the fixed parameters, gamma priors for the scalar variances, and inverse-Wishart priors for the variance matrices.

Missing data

Missing data is a problem in all trials. When data is missing it has the potential to reduce the overall power of the study to detect a small effect size. Worse still, when data is 'missing not at random' this has the potential to bias the reported effect. Multilevel multiple imputation using the beta version of the new n-level software package, STAT-JR, was not used as originally planned because the software did not run. Instead we used the missing data features within STAN to impute values. This entailed a deviation from the published SAP, however given that it is based on the same theoretical approach to imputation (joint modelling), the sampling process accounts for uncertainty and the approach used the same auxiliary variables to extend the Missing At Random assumption, it is very unlikely to have any important implications for the analysis.

Compliance

A set of criteria for school-level compliance was agreed with the developer that covered attendance at training, adequate staffing, initial assessment, termly review, pupil selection, and the delivery of support sessions in school (see Appendix H). The original intention, as described in the SAP, was to create a three-fold category of high, medium, and low fidelity at a school level, based on the criteria set out in Appendix H, then to conduct per-protocol analyses based on two levels of fidelity: high, and high and medium combined. However, as is described below, it was not possible to fully implement the criteria in Appendix H because some data was unavailable. Moreover, very few schools met the criteria for pupil selection, whilst almost all schools met the criteria on attendance at training, delivery of support sessions, termly review, and initial assessment. Hence, an alternative approach was adopted involving descriptive analysis for all aspects except pupil selection.

Attendance at training and adequate staffing

Developer records were available for attendance at training, which was very high (at 97% attendance). However, records were unfortunately not consistently available for staffing, because many schools did not inform the developer of staff changes.

Delivery of support sessions, termly review, and initial assessment

The developer also collected records about the delivery of support sessions as well as termly review and roll-off. However, using these to assess the effects of non-compliance is not straightforward. First, there was a large amount of missing pupil data in the developer records: although 68 (91%) of schools submitted records of support sessions, the data was complete for all pupils in only 20 (27%) of intervention schools. Second, some schools recorded the assessment sessions whereas others did not. Third, the effects of support are affected by two further factors: dosage and composition.

By design, the intervention involves different levels of dosage depending on when pupils are rolled off the intervention. Unfortunately, although data was collected from active control schools on dosage (that is, the length as well as the number of sessions), similar data was not available for intervention schools because the developer collected only data on the number of sessions, not the length of sessions. Data was collected on dosage in the TA survey. The effects of dosage were investigated at TA level descriptively using a four-category item from the TA survey.

Composition—whether support was delivered individually or in groups—is also likely to affect the impact of the support delivered. Data was collected on composition in the TA survey and the effects were

investigated at pupil-level by including a five-category item from the TA survey within the regression model. However, as is discussed in the analysis section, for the intervention group, compliance to the 1:1 delivery was very high (98%), although this was not the case for the active control.

Pupil selection

As already noted, the schools in this trial had limited guidance on how to select pupils and could not follow the usual Catch Up® Numeracy practice of using a standardised test as part of the selection process. As a result, assessing compliance to pupil selection is not straightforward. However, the Catch Up® Numeracy intervention is aimed broadly at the lowest attaining 15–20% of pupils based on their age. Since pupils were selected by schools, it was judged likely that some middle- or high-attaining pupils outside this group would be selected, and, although pupils were selected for the intervention prior to randomisation, it is possible that such mis-selection would result in some bias. Specifically, support provided in the active control could have been more appropriate for pupils outside the target group than Catch Up®. In order to test this, the interaction between prior attainment and treatment group was investigated. In addition, the effects of compliance with pupil selection were also investigated using a per protocol analysis at pupil level and Key Stage 1 (KS1) average points score data, based on a comparison of eligible pupils (with the lowest attaining 15–20% pupils criterion operationalised as those with a KS1 score less than 15).

In summary, attendance at training, initial assessment, termly review, dosage, and composition were investigated descriptively. Pupil selection was investigated quantitatively using a per protocol analysis.

Implementation and process evaluation

Data collection

The process evaluation involved multiple methods to capture the overall views of all participants in the trial using surveys complemented by interviews with key personnel and in-depth case studies of selected intervention and control schools.

Data on attendance at training sessions and records of support sessions (dosage) and initial assessments and termly assessments (or roll-off) were provided for the Catch Up® schools by the developer. All other implementation, process, and fidelity data was collected and analysed by the evaluation team.

Training session surveys

Before the intervention began, surveys of their experiences at the three initial Catch Up® Numeracy training events for TA and school coordinators were provided to all attendees in order to address RQ6 (effectiveness of training). These asked for the perceptions of the quality of the training as well as their initial views on Catch Up® and their confidence to implement it. The 207 attendees were given surveys, of which 97% were returned. This data was complemented by observations of four professional development events. After each of these events, the trainer and two participants were interviewed.

Developer records of attendance at training

The developer provided records of attendance at training which were used to address RQ6 (effectiveness of training).

Developer interviews

Two interviews with the Catch Up® team of developers were conducted to explore the planning, design, and logistics behind the intervention. These interviews were used, alongside other resources and

training materials provided by the developer, to inform the data collection and analysis and are not analysed directly in this report.

Surveys for all the teaching assistants, coordinators, and senior leaders in Catch Up® schools

TA surveys primarily addressed RQ6 (perceptions of the effectiveness of training) and RQ7 (adherence, fidelity, quality, and implementation factors) concerned perceptions and experiences of Catch Up® and its impact upon pupils, fidelity to the intervention protocol, and whether their experience of Catch Up® had influenced their attitudes to mathematics and their teaching of it. The surveys for coordinators and senior leaders were designed to address RQ6 and RQ7 (in particular schools' perceptions of intervention effectiveness and factors relating to implementation). The surveys asked similar questions to each other. These surveys asked questions concerning fidelity that were pertinent to their role, as well as their perceptions of the impact of Catch Up® Numeracy upon the TA and class teachers, in addition to pupils. These surveys were distributed to all participants towards the end of the school year: 146 to TAs and 73 (i.e., one per school) to coordinators and senior leaders. The return rate was 75%, 84% and 68% respectively.

Surveys for all the teaching assistants, coordinators, and senior leaders in active control schools

Equivalent surveys for TA, TA coordinator, and senior leader participants in active control schools were designed to address RQ9 (monitoring of control group and usual practice). These surveys were developed following the same format. However, specific references to Catch Up® were replaced with more general terminology referring to the active control intervention. The number of surveys distributed to each sample were 149 to TAs and 75 to coordinators and senior leaders. This resulted in a 73%, 85% and 65% return rate respectively.

Surveys of teaching assistant attitudes to mathematics and mathematics teaching

Surveys of TA attitudes were developed following a similar model to the pupil attitudes survey described above. These were administered at the start of the trial, but post-randomisation. However, initial analysis indicated a ceiling effect: almost all the TAs reported that they were 'confident' or 'very confident' in numeracy and the teaching of numeracy. [6] As a result, this data was not used in the analysis, and retrospective Likert-scale items were included in the TA survey (such as, 'As a result of delivering the intervention I like maths more').

Logs of support provided for pupils in active control schools

In order to address RQ5 (dosage) and RQ8 (the effects of roll-off), active control schools were asked to complete a half-termly record of support sessions delivered for each pupil (including the length of the sessions) and whether pupils were rolled-off the programme as a result of regular termly reviews. Logs were sent to 75 active control schools and the return rate for the six half-terms was 92%, 76%, 76%, 73%, 73% and 47%, respectively.

Developer logs of support for pupils in Catch Up® schools

Similar records were required from Catch Up® schools in order to address RQ5 and RQ8. The developer provided data based on its normal record-keeping on the number of support sessions delivered for each pupil, the initial assessment, and whether pupils were rolled-off the programme. These records were provided by 89% of Catch Up® schools, although records were missing for some additional pupils. Unfortunately, the Catch Up logs recorded only the number of sessions and not the length of sessions, and there was significant missing data, meaning the comparisons of dosage planned in the SAP were

[6] The tests were administered to the intervention TAs during their initial training. The ceiling effect was immediately apparent. As a result, the TAs in the control group were not asked to complete the surveys.

not possible using this data. However, a four-category item on the TA survey asked about length of sessions[7] and was used instead together with case study data.

Case studies of Catch Up® and active control schools

The case study aspect of the process evaluation was designed to examine in greater depth how Catch Up® (CU) was implemented in the intervention schools and how the active control (AC) group supported numeracy for their pupils. It involved two visits to intervention and one to control schools to conduct in-depth interviews with TAs, teachers, coordinators, and senior leaders. All visits involved observation of multiple numeracy support sessions and inspection of pupil work.

The focus of the first visit concerned existing practice in the schools and staff's initial perceptions and experiences of Catch Up®. The second visit permitted additional reflection towards the end of the intervention on its outcomes, as well as the opportunity to interview participants not available on the first visit. Both visits included observations of support sessions. Only a single visit to the control group schools was required as this was sufficient to capture staff experience of their existing practice as well as the extent to which involvement in the active control arm trial had changed previous practice.

Schools were selected in the following way. All schools were asked if they would be prepared to be involved in the case study. Of those that responded positively, five intervention case study schools were selected to ensure that data could be generated from schools implementing Catch Up® as close to the intended approach and dosage as it was possible to determine from the records and then to maximise variability in the economic profile of the school's catchment area. In addition, a contrasting school was selected as it reported less commitment to Catch Up® (as evidenced by completion of session logs). Control schools were selected to maximise the variability in the school's catchment area.

The five Catch Up® Numeracy case-study schools are as follows.

- *School A (CU)* is a larger than average local authority primary school situated in a suburban area of large city. It has a below average proportion of pupils eligible for free school meals (FSM) but a greater than average proportion of students with English as an Additional Language (EAL) and with Special Educational Needs (SEN). In this school, a member of the school senior leadership team coordinated the intervention TAs, so there was no separate TA coordinator. Two TAs, a teacher, and the coordinator/senior leader were interviewed.
- *School B (CU)* is a larger than average academy-sponsored primary school situated in a small town. It has a greater than average proportion of pupils eligible for FSM as well as those with SEN but a lower average proportion of EAL students. Two TAs, two teachers, and the coordinator were interviewed.
- *School C (CU)* is an academy converter primary school situated in medium sized town. It has a below average number of children eligible for FSM and with a SEN statement but an above average number of children with EAL. Two TAs, two teachers, the TA coordinator, and a senior leader were interviewed.
- *School D (CU)* is a much larger than average local authority maintained primary school situated in a medium sized town. It has an average number of children who speak English as an Additional Language and below average number of FSM and SEN pupils. Two TAs, a teacher, the TA coordinator, and a senior leader were interviewed.
- *School E (CU)* was selected as the contrast school as it had decided to withdraw from Catch Up. It is a local authority primary school situated in a large town and is larger than average. It has a lower than average proportion of FSM, SEN and EAL pupils. Two TAs, two teachers, the TA coordinator, and a senior leader were interviewed in a single visit.

[7] One question was, 'How long do most [Catch Up®/numeracy] sessions last?', with four options: 'over 15 minutes', '15 minutes', 'under 15 minutes', and 'varies from week to week'.

The five active control schools are as follows.

- *School A (AC)* is a larger than average recent academy converter primary school situated in Greater London. It has a higher than average number of children who speak English as an Additional Language, are eligible for free school meals and have statement of special educational needs. Two TAs, a teacher, the TA coordinator, and a senior leader were interviewed.
- *School B (AC)* is a larger than average local authority maintained primary school situated in a large town. It has a greater than average number of FSM pupils. It has a lower than average number of children who speak English as an Additional Language, or have a statement of special educational needs. Two TAs, a teacher, the TA coordinator, and a senior leader were interviewed.
- *School C (AC)* is a larger than average academy converter school in medium sized city. The majority of pupils speak English as an Additional Language. There is a higher proportion than average eligible for free school meals and a lower than average proportion have support for special educational needs. Two TAs, two teachers, the TA coordinator, and a senior leader were interviewed.
- *School D (AC)* is a smaller than average academy school in a small town. The proportion of FSM, EAL and SEN pupils are all below average. Two TAs, two teachers, and the TA coordinator were interviewed. Uniquely, it was not possible to conduct observations in this school.
- *School E (AC)* is a larger than average recent academy converter school in a large town. It has a higher than average number of children who speak EAL and are eligible for FSM and an average with a statement of SEN. Two TAs, two teachers, and the TA coordinator were interviewed.

Data analysis

Survey data

The survey data was analysed descriptively and using inferential statistics. For the purposes of the IPE, multiple t-tests were used to investigate differences between the means of Likert scale items for the Catch Up® and active control groups (using Welch's test for unequal variances). Any differences were hypothesised to be relatively small and, in order to avoid false negatives, no correction was made for multiple testing (MacDonald, 2014). In addition, no account was made for any clustering of the sample (TAs within schools). Hence, the results of this analysis should be treated with some caution and as indicative rather than conclusive.

Case study data

The case study data (interviews and observations) were analysed using thematic analysis (following, for example, Braun and Clarke, 2006) to answer the research questions highlighted earlier, explore the dimensions and factors affecting implementation (Humphrey et al., 2016), and address the set of ten questions highlighted in the EEF evaluation report template (EEF, 2018). An initial set of codes was generated in two ways. Firstly, two earlier reports were scrutinised. The first was a previous evaluation report of an efficacy trial of Catch Up® Numeracy (Rutt et al., 2014), which was read to identify factors previous noted as influencing the implementation of Catch Up®. The second was a report on the effective use of teaching assistants (Sharples, Webster and Blatchford, 2015) so as to identify known issues that could apply to the control schools as well as the Catch Up® schools. At the same time, all transcripts were read to inductively identify codes. This resulted in an initial set of 114 codes. Then over a period of review and refinement by three of the authors of this report, a set of themes and subthemes were ultimately generated, which form the basis for the report of the analysis.

Fidelity

Data from surveys, logs of support provided, and case studies provided evidence relating to fidelity and compliance relating to dosage, composition (group or individual support), and roll-off. Test data collected for the impact evaluation (including Key Stage 1 scores) provided evidence about adherence to the selection of pupils.

Costs

Cost information was based on the commercial costs of the training provided by the developer, and on the assumption that three staff members would be trained as in this trial (two TAs and one school coordinator). The travel to regional training for some schools was relatively substantial. Hence, in addition, the travel costs required to attend the regional training were estimated. This based on the average distance from schools (41 miles), assuming three people car-sharing and using the University of Nottingham mileage rate of 40p per mile. Per-pupil costs were calculated on the assumption of 12 pupils per school (and six pupils per TA) to determine the cost per pupil per year.

During the case studies carried out for the implementation and process evaluation, questions were asked about additional resources and workload required to implement the Catch Up® intervention.

Timeline

Table 5 sets out a timeline of the main activities carried out by the developer and the evaluation team.

Table 5: Timeline

Date	Activity
January–September 2016	Recruitment of participating schools (led by developer)
June-July and September 2016	Selection of pupils, and appointment of staff, by schools
June–July and September 2016	Pre-test administration by schools
July and September 2016	Randomisation (after pupil selection and pre-test administration)
July and September 2016	Training for TAs and school coordinators
September 2016–July 2017	Implementation in schools
June and July 2017	Surveys of TAs, coordinators, and SLT
February–June 2017	Case study interviews and observations
October–December 2017	Delayed post-tests (administered by evaluation team to intervention and active control schools)

Impact evaluation

Participants

Figure 3 outlines the flow of participants during the course of the project. One hundred and fifty-one schools were randomised into the Catch Up® Numeracy and active control arms, with one Catch Up® school withdrawing before notification of the treatment assignment, two further Catch Up® schools shortly after the start of the main trial in September 2016, and a further Catch Up® in Summer 2017. A further three schools—one in the Catch Up® arm and two in the active control—were lost to the follow-up post-test in October–December 2017. The intervention school had fully implemented the Catch Up® Numeracy intervention, but was unwilling to administer additional tests to their pupils.

Although attrition was within acceptable bounds at the school level (5%), at pupil-level, attrition was high (17%). Of the 1,794 pupils who received treatment in the Catch Up® and active control arms at the start of the main trial, 1,775 were matched to National Pupil Database records (after the withdrawal of the Catch Up® schools) and complete records for the primary model were available for a total of 1,481 pupils. This was 83% of pupils randomised—737 pupils receiving Catch Up® and 744 pupils receiving the active control condition.

Attrition was due to a combination of factors in addition to school and pupil withdrawals:

- Delayed post-testing: the delayed post-testing took place in the autumn term of the academic year following implementation. As a result, 129 pupils, 7% of pupils allocated, had left school (including transfers to middle schools).
- Pupil absence: 107 pupils were absent at the time of post-testing; 49 pupils completed the post-tests later. However, a total of 58 pupils, 3% of pupils allocated, were missing post-tests.
- Missing pre- and post-tests: despite considerable efforts and additional expense by the evaluation team, GL Assessment were unable to account for 60 pre-tests and 7 post-tests, leading to further attrition of 49 pupils (42 pre-tests and 7 post-tests). This constituted 3% of pupils allocated and resulted in the loss of pre-tests for two complete schools.

Without the attrition due to pupils leaving schools and the missing GL tests, the level of attrition would have been well below 10%.

Statistical power

The Minimum Detectable Effect Size (MDES) was calculated using Optimal Design at all stages, on the basis of a three-level cluster random design (pupils nested in TAs, and further nested in schools) with treatment at the school level. The power calculations assumed six pupils would be taught by each TA, and two TAs would be employed per school focused on the delivering numeracy intervention. Inter-Class Correlations were calculated for both the TA level and school level.

Table 7 shows the MDES at each stage in the development of the trial. It also shows estimates of the correlation between pre-test and post-tests, and of intra-cluster correlations. Specifically, it highlights that the combination of a weaker relationship between the pre-test and post-test than predicted, and the presence of significantly more variation than had been hypothesised for each level of the analysis, raised the MDES from 0.17 to 0.25 and 0.31 for the entire sample and the sub-sample of FSM students, respectively. The ICCs are very much higher than those used for the original power calculations in the protocol and are surprising given the ICC levels in other EEF trials (see Table 6 and Table 7). This can be explained by the presence of floor effects within the pre- and post-tests (and poor correlations between PTM and KS1 scores), and the mis-selection of pupils. The distribution of the tests and the pupil selection issue are discussed further below.

Table 6: Intra-cluster correlation analysis

Test	Intra Cluster Correlation		
	TA level	School level	Region level
Post-test (Protocol)	0.05	0.10	-
Post-test (Analysis)	0.09	0.25	0.06
Post-test (Imputed)	0.14	0.23	0.11
Pre-test	0.14	0.23	0.11
KS1 Average Points Score	0.10	0.22	0.05

Figure 3: Participant flow diagram

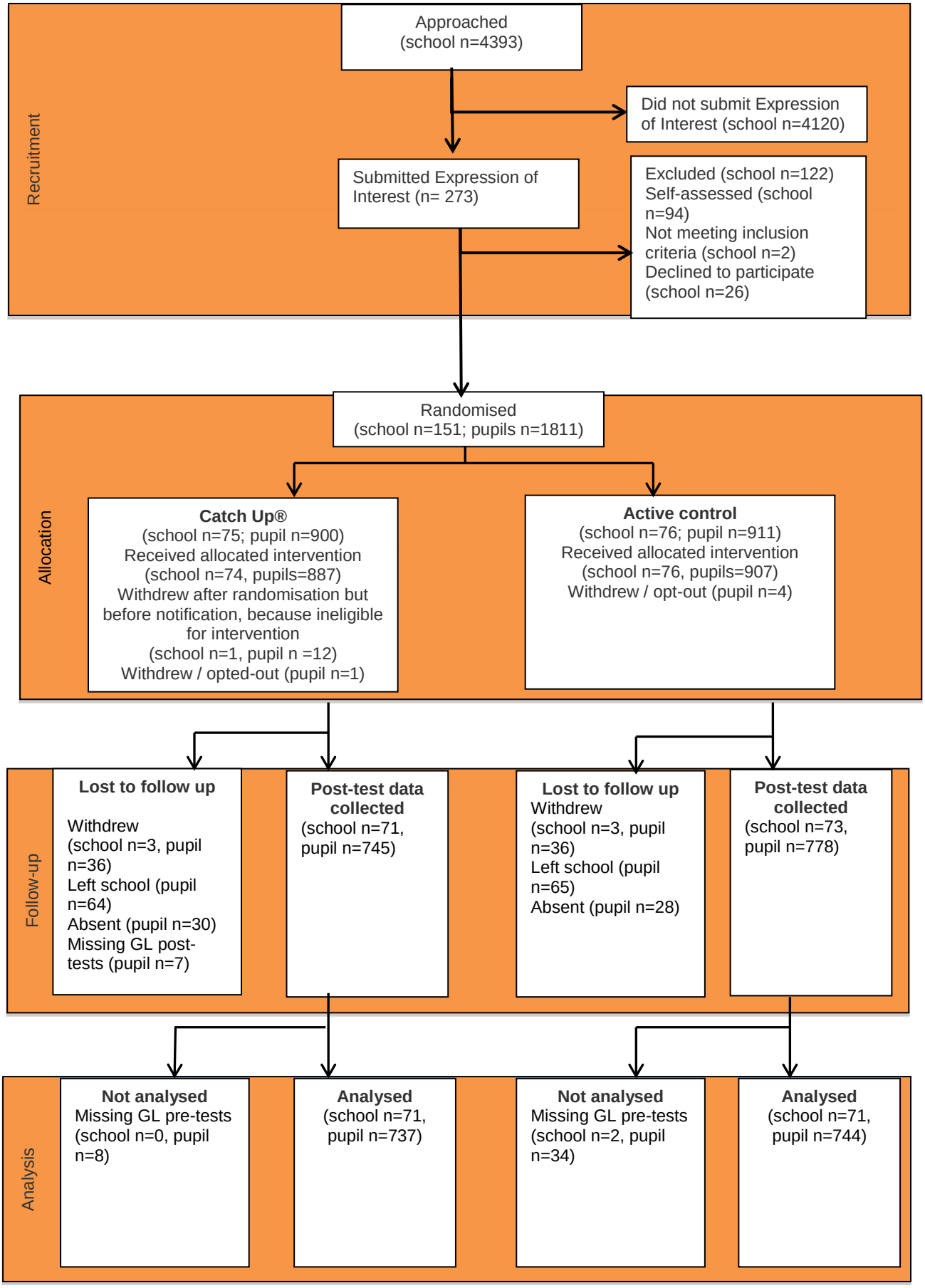


Table 7: Minimum detectable effect size at different stages

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.17	0.25	0.17	0.25	0.25	0.31
Pre-test/ post-test correlations	level 1 (pupil)	-	-	-	-	0.69	0.67
	level 2 (class)	-	-	-	-	-	-
	level 3 (school)	0.8	0.8	0.8	0.8	-	-
Intracluster correlations (ICCs)	level 2 (class)	0.05	0.05	0.05	0.05	0.09	0.06
	level 3 (school)	0.10	0.10	0.10	0.10	0.25	0.28
	level 4 (Region)	-	-	-	-	0.06	0.04
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided
Average cluster size		12 pupils	12 pupils	4 pupils	12 pupils	10 Pupils	4 Pupils
Number of schools	intervention	75	75	75	75	71	66
	control	75	75	76	76	71	66
	total	150	150	151	151	142	132
Number of pupils	intervention	900	300	900	300	737	290
	control	900	300	911	304	744	261
	total	1800	600	1811	604	1481	551

Pupil and school characteristics

Table 8 examines the school and pupil-level baseline characteristics and balance on key variables. The first point to note is that there was a significant imbalance in the pre-test PTM scores at the pupil level between the two arms (effect size = 0.21), which is at a level that has implications for the robustness of the findings. This imbalance was unexpected and is surprising given the much smaller level of imbalance in Key Stage 1 scores (effect size = 0.06), as well as the low level of imbalance at randomisation. It is important to note also that at randomisation the imbalance was even smaller (effect size = 0.03) but due to attrition this rose to 0.06.

One reason for this imbalance may be due to the GL PTM tests themselves. In particular, the correlation between Key Stage 1 scores and the PTM tests was low: for the age-standardised scores it was 0.48 for the pre-test and 0.43 for the post-test; for the raw scores with the Year 4 group it was 0.48 for the pre-test and 0.50 for the post-test; and with Year 5 it was 0.51 for the pre-test and 0.45 for the post-test. In addition, there were substantial floor effects on the PTM age-adjusted scores. This is discussed further in the outcomes section below and additional results are presented in Appendix E. A second reason may be due to the selection of pupils. As already noted, the selection of pupils was carried out prior to randomisation and before training in order that selection was not influenced by group allocation. As is discussed in the dosage and compliance section below, a large proportion of the pupils selected

by schools were middle and high attaining (for example, those who attained an average score of 15 or more at KS1 was 36%). Indeed, some schools selected pupils who were almost all middle and high attaining; two control schools and four Catch Up® schools selected ten or more pupils in the middle to high attaining range.

At the pupil level, there were also smaller imbalances in the percentage of pupils eligible for FSM and everFSM, with Catch Up® schools having 4% more than control schools. The groups were balanced on gender with 51% of both groups being female pupils.

On the school-level categorical measures, the numbers of LA maintained and academy-run schools are very similar between the Catch Up® and active control groups. While there are clearly more Ofsted-rated 'outstanding' schools (6% vs. 15%) within the control group, when this is broadened to examine both schools rated 'outstanding' and 'good' combined, there is strong evidence of balance between the groups (87% vs. 86%). The groups are very similar in terms of their general location, with 80% of Catch Up® schools based in urban areas vs. 84% of active control schools.

For the school-level continuous measures, there is evidence of some imbalance present in both the average percentage of everFSM pupils—with Catch Up® schools having on average approximately 4% more pupils from deprived backgrounds—and furthermore, an imbalance on the percentage achieving level 4 at KS2 of approximately 2.2% with Catch Up® schools performing better, on average, than active control schools. Adjustments were made to the model in the sensitivity analysis to adjust for the imbalance present in the school-level continuous measures and these are discussed below.

Table 8: Baseline comparison at analysis stage (based on complete case analysis of primary model parameters) —1,481 pupils in 71 Catch Up® Numeracy schools and 71 active control schools

School-level (categorical)	Intervention group		Control group		
	n/N (missing)	Percentage	n/N (missing)	Percentage	Effect Size
School type: LEA Maintained Academy	71/71 (0) 45/71 26/71	100% 63 % 37 %	71/71 (0) 43/71 28/71	100% 61% 40%	N/A
Ofsted rating: Outstanding Good Requires Improvement Inadequate	71/71 (0) 4/71 58/71 8/71 1/71	100% 6% 82% 11% 1%	70/71 (1) 11/71 49/71 9/71 1/71	99% 15% 70% 12% 1%	N/A
Urban vs. rural location: Urban Rural	71/71 (0) 56 15	100% 80% 20%	71/71 (0) 60 11	100% 84% 16%	N/A
School-level (continuous)	**n (missing)**	**Mean (SD)**	**n (missing)**	**Mean (SD)**	**Effect Size**
Percentage eligible for FSM (in past 6 years)	71/71 (0)	30.5 (16.7)	71/71 (0)	26.5 (16.3)	N/A
Percentage achieving level 4 in KS2 combined scores	69/71 (2)	78.4 (13.1)	69/71 (2)	76.18 (15.9)	N/A
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**	**Effect Size**
Eligible for FSM	736/737 (1)	24%	744/744 (0)	20%	N/A
Eligible for FSM (in past 6 years)	730/737 (7)	39%	735/744 (9)	35%	N/A
Gender	737/737 (0)	51% (Female)	744/744 (0)	51% (Female)	N/A
Pupil-level (continuous)	**n (missing)**	**Mean (SD)**	**n (missing)**	**Mean (SD)**	**Effect Size**
Age standardised score on PTM	737/737 (0)	86.1 (9.7)	744/744 (0)	84.0 (9.6)	0.21
KS1 Average Points Score	737 (45)	13.6 (2.4)	744 (31)	13.8 (2.4)	0.06

Outcomes and analysis

Primary outcome

The primary analysis of the impact of Catch Up® Numeracy against the matched-time active control arm was assessed on the basis of intention-to-treat using a multilevel linear regression model estimated by Bayesian inference. The primary outcome of interest—the average difference in the post-test scores between the Catch Up® Numeracy treatment group and the active control condition group—was measured by using the standardised post-test score on the Progress Test in Mathematics provided by GL Assessment. The equation below is as set out in the statistical analysis plan, with one minor modification in that randomisation group was included as a separate grouping parameter:

$$y_{ijkl} = \beta_0 + \beta_1 Treatment_i + \beta_2 Pre-test_i + \underbrace{f_{0l} + v_{0k} + u_{0j}}_{\text{varying intercepts}} + \epsilon_{ijk}$$

$$\begin{aligned} f_{0l} &\sim N(0, \sigma^2_{Randomisation\ Group}, for\ l = 1 \ldots L \\ v_{0k} &\sim N(0, \sigma^2_{School}, for\ k = 1 \ldots K \\ u_{0j} &\sim N(0, \sigma^2_{Teaching\ Assistant}, for\ j = 1 \ldots J \\ \epsilon_i &\sim N(0, \sigma^2_{\epsilon}, for\ i = 1 \ldots N \end{aligned}$$

The individual level of the model has a grand mean of the PTM post-test (represented by β_0) which varies by membership of TA group, School, and Randomisation Group (represented by the intercept adjustments u_{0j}, v_{0k} and f_{0l}), an individual-level binary treatment covariate (where 0 represents those pupils who received the matched-time intervention and 1 which represents those pupils who received the Catch Up® Numeracy intervention), a normally distributed and mean-centred pre-test covariate, and lastly an error term (ε_{ijk}).

Effect sizes distributions were simulated using the following approach. The fitted model posterior samples were extracted in order to compute effect sizes quantities of interest for all relevant parameters which allowed us to generate means, standard deviations, and 95% credible intervals. In particular, effect sizes were calculated using total variance as set out in the following formula:

$$ES = \frac{\bar{Y}_t - \bar{Y}_c}{\sqrt{(\sigma^2_{randomisation\ group} + \sigma^2_{school} + \sigma^2_{ta} + \sigma^2_{y})}}$$

Table 9 summarises the raw means of the intervention and control group from the entire dataset across both primary and secondary outcome measures—the Progress Test in Mathematics post-test and the Attitudes to Mathematics post-test conducted as a delayed follow-up in October–December 2017. It also presents the impact of the final model on the number cases and the computed effect sizes. The impact on the primary and secondary outcomes was assessed through the use of Bayesian multilevel modelling. It should be noted that statistical significance and *p*-values as classically presented do not exist within the Bayesian statistical paradigm. In each case, a 95% Bayesian credible interval and a Region of Practical Equivalence (ROPE) have been calculated to aid interpretation. Specifically, as described in the methods, the ROPE reports the proportion of the highest posterior density interval (HDI) that falls ± 0.1 around 0, and even effects that are bounded away from 0, but fall predominantly within the ROPE, are in effect a credible or practically insignificant null effect. Classical analyses were also conducted and can be found in Appendix E, Tables 6–9. The results were consistent with the Bayesian analyses.

The raw means on the post-test scores of the PTM test suggest that the Catch Up® group scored on average 1.4 marks higher than the active control group, although with the intervals overlapping this was

not a clear effect. When modelled to take account of the prior attainment measure—the score on the PTM pre-test—the positive difference in the means was reversed, and when standardised this resulted in an effect size of -0.04 (-0.21, 0.13).[8] The additional ROPE analysis suggests that 72% of the 95% Bayesian highest posterior density interval (HDI) falls within ± 0.1 around 0. As such, having updated the vague existing knowledge with new data, this indicates that there is no impact of Catch Up® against active, matched-time equivalent.

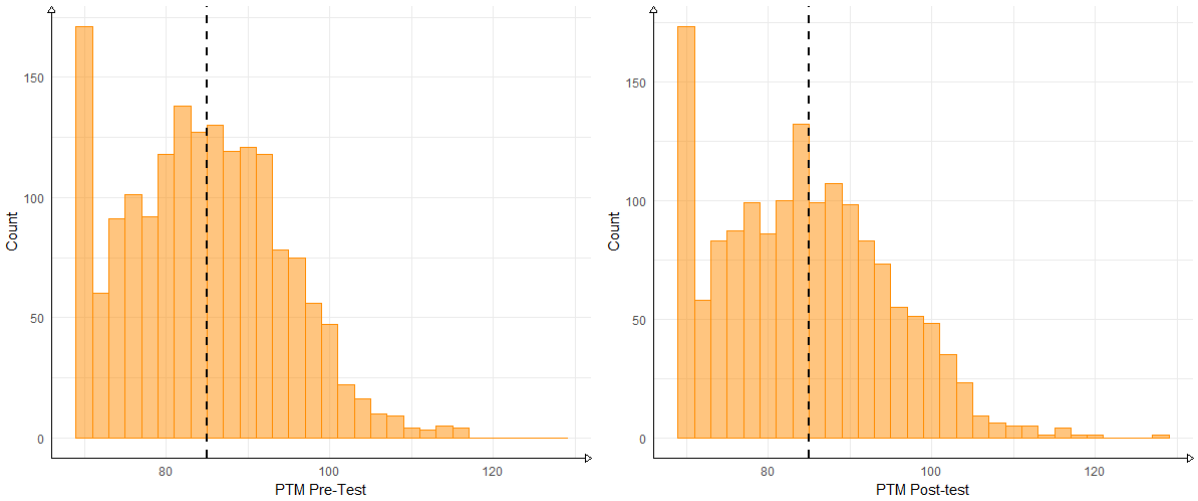
Table 9: Summary of mean differences and effect size of the outcome measures

	Raw means				Effect size		
	Catch Up® group		Active Control group				
Outcome	n (missing)	Mean (95% Confidence Interval)	n (missing)	Mean (95% Confidence Interval)	n in model (intervention, control, missing)	Bayesian Hedges g (95% Credible Interval)	Region of Practical Equivalence
Post-test PTM Age adjusted score	876 (131)	85.7 (85.0, 86.5)	899 (121)	84.3 (83.6, 85.0)	1481 (737, 744, 318)	-0.04 (-0.21, 0.13)	72%
Post-Test Attitudes	876 (262)	7.7 (7.6, 7.8)	899 (260)	7.5 (7.4, 7.7)	1124 (547, 577, 675)	0.11 (-0.05, 0.27)	44%

Distribution of the PTM tests

As noted above, although GL Assessment advised that the relevant PTM tests were appropriate for low attaining students, previous trials suggest that the use of age-standardised scores may in certain cases inflate effect sizes due to floor and ceiling effects. Figure 5 illustrates the distribution of test scores for the primary outcome analysis using all available cases.

Figure 5: Histograms of the Progress Test in Mathematics pre- and post-test scores



In both the pre-test taken in June 2016 and the post-test taken in October–November 2017, there were significant floor effects present. On the post-test, 70 (4%) scored the lowest score of 69, and 173 (10%) scored less or equal than 71. On the pre-test, 119 (7%) pupils scored the 69 and 171 (10%) scored less than or equal to 71. These floor effects were not present in the raw scores and arose through the age

[8] When modelled as originally intended in the SAP using three-level linear multilevel regression (ignoring randomisation block), the standardised effect size was also found to be -0.04.

standardisation process. The distribution of KS1 Average Point Score and PTM pre-test by treatment arm is included in Appendix E, Figures 3 and 4. The raw PTM scores could not be used as an outcome measure because of the need to equate the scores for pupils of different ages. However, analysis of the effects within Year 4 and Year 5 indicated no difference between treatment arms in the different age groups (see Appendix E, Table 1).[9] Given the focus of the intervention on a subset of low attaining pupils, this created a significant problem in that the floor effects had the potential to significantly dilute estimates of the impact of the intervention. Hence, as planned, a Bayesian censored student-t multilevel regression model (an equivalent of a Tobit model) was fitted to allow a linear model to be fitted when the data is censored by floor or ceiling effects, and to check the robustness of the regression findings given the observation of 'fatter tails' of the distribution. The results were consistent with the primary model, although the estimated variance increased slightly at all levels (the unstandardized regression results are presented in Appendix E, Table 12).

Secondary outcome

The secondary outcome analysis of the impact of Catch Up® Numeracy on children's attitudes towards mathematics when compared to an active matched time control, as with the primary outcome, was assessed on the basis of intention-to-treat using a multilevel linear regression model estimated by Bayesian inference. The outcome of interest was the average difference in the post-test scores on a bespoke pupil attitudes survey between the Catch Up® Numeracy treatment group and the active control condition group.

The raw means suggest that the Catch Up® group averaged 0.2 points higher on the post attitudes survey (see Table 9). When adjusted for pupil scores on the pre-test, the active control group averaged 7.51 points (7.15, 7.89), with the Catch Up® group scoring 0.17 points higher (-0.07, 0.41), representing a standardised effect size of 0.11 (-0.05, 0.27). As the Bayesian credible intervals highlight, this effect is not bounded away from 0. Nevertheless, given the substantial proportion of the posterior distribution being greater than 0, this result is worthy of further examination. The ROPE analysis suggests that 44% of the 95% Bayesian HDI falls within ± 0.1 around 0. Given that the HDI's lower bound was less than -0.1, 66% of the interval falls above +0.1, confirming that there is weak evidence of a positive effect.

Subgroup analyses

Two subgroup analyses were carried out examining the effect of Catch Up® Numeracy on subgroups relating to gender and FSM status using an interaction effects analysis. For boys, the effect of Catch Up® was -0.06 (-0.21, 0.17) and for girls, the interaction effect was g -0.01 (-0.21, 0.17). For everFSM pupils the effect was -0.14 (-0.33, 0.05) and for non-FSM it was g = 0.02 (-0.16, 0.21). With the exception of the everFSM pupils, all effects were centred around 0 suggesting that on the basis of the available evidence, Catch Up® Numeracy does not have a clear effect over the matched time active control.

For the treatment effect on everFSM pupils, while it crossed the 0 boundary, 92% of the posterior distribution was negative giving stronger evidence in this case of Catch Up® treated pupils performing, on average, more poorly than their active control counterparts. A further analysis on the FSM subset produced a similar effect with the difference between the two due to the impact of the prior and different variance parameter values (see Appendix E, Tables 15 and 16). To examine further the robustness of this effect on FSM pupils, we compared the results of modelling the raw PTM post-test scores with the PTM pre-test as a covariate, with KS1 scores as a covariate and with both as covariates (using the sample of pupils for which both PTM pre-test and KS1 scores were available). All three models produced a similar effect (see Appendix E, Table 17). However, some caution must be exercised in that

[9] The sample of Year 3 students was considered too small to carry out a sub-group analysis at this age.

these analyses involve small samples and should be the focus of further research before stronger conclusions can be drawn.

To adjust for the imbalances present in the sample, additional analyses were carried out using the following school-level covariates which were mean centred: average KS1 score and average KS2 score. Despite the additional adjustments, the results remained consistent with the primary model, with a main effect centred on 0 with the credible interval crossing 0 (see Appendix E, Table 13).

Sensitivity analysis

There were four main areas of attention for sensitivity analyses which were set out in the statistical analysis plan. One of these, the examination of the implications of floor effects on the test scores is discussed above (with only a minor deviation from the SAP by adjusting for the 'fatter' tails of the distribution). The remaining analyses were using alternative prior distributions, estimating the primary model using classical (rather than Bayesian) methods, fitting the models using alternative software, fitting the models with school-level covariates, and fitting a multivariate multilevel model. This introduced an additional parameter which identified the censored observations and adjusted the standard deviations and variance parameters accordingly. The primary model was refitted using MLwiN's IGLS (which is a maximum likelihood based algorithm) as well as with a Gibbs Sampler (a Markov Chain Monte Carlo sampling algorithm) using the standard diffuse priors. The results from IGLS and MCMC remained consistent with the results from Stan and lme4 in R, although the Gibbs sampler took significantly longer to converge to a stable distribution. The classical, censored student-t regression and primary model with school-level covariate results are reported in the Appendix E, Tables 8, 9, 12 and 13 respectively, and the multivariate multilevel model results are presented in Appendix F, Table 1.

Missing data

Missing data remained a relatively small, but important issue to address. Table 10 sets out the amount of missing data in pupil records for the individual level and school-level covariates. Fourteen percent of pupil records had missing post-test data and 3% were missing the pre-test. Broken down by treatment group, 15% of Catch Up pupils were missing post-tests compared to 13% of active control pupils. On the pre-test, 2% of Catch Up pupils and 5% of active control pupils had missing records. Turning to missingness within the secondary analysis and potential auxiliary variable covariates, only 1% of pupils were missing everFSM data, 5% were missing pupil KS1 data, 6% were missing school-level KS1 records, and 3% were missing school-level KS2 records.

Table 10: Missingness in the outcome, pupil-level, and school-level covariates overall and by treatment group

Variable	Missing (%)	Missing Catch Up (%)	Missing active control (%)
Post-test	252 (14%)	131 (15%)	121 (13%)
Pre-test	60 (3%)	18 (2%)	42 (5%)
everFSM	20 (1%)	9 (1%)	11 (1%)
Pupil KS1	91 (5%)	49 (6%)	42 (5%)
School KS1	108 pupils in 9 schools (6%)	48 pupils in 4 schools (5%)	60 pupils in 5 schools (7%)
School KS2	48 pupils in 4 schools (3%)	24 pupils in 2 schools (3%)	24 pupils in 2 schools (3%)
Total	1775 pupils in 148 schools	876 pupils in 73 schools	899 pupils in 75 schools

A dropout model of missingness was constructed for the primary model where each case that was completely observed was coded with a 0 and each case that was either missing data on the post-test was coded with a 1. This variable was specified as an outcome in a multilevel logistic regression model and regressed against the treatment condition, pre-test, everFSM status, pupil-level KS1 score, and the school-level KS1 and KS2 averages. The model also included the indicators for TA assignment, School, and Randomisation group. Table 10 in Appendix E presents the findings from the model. Most predictors were not bound away from 0. However, pupil KS1 scores and everFSM did, and when converted into probabilities, a one unit move for pupil KS1 score decreased the probability of missingness of an average pupil from 4.6% to 4.2%, and a one unit move for everFSM increased the probability of missingness from 4.2% to 9.8%.

Originally in the SAP the missing data approach was to impute the primary model missing results using a new n-level multilevel multiple imputation model. However, due to teething difficulties in running the software, and the confinement of missing values to continuously distributed variables, the missing values were imputed within STAN itself. Bayesian methods treat missing data as a random parameter which can be explicitly modelled and estimated simultaneously with the main model. To this end, several sub-models were added to the primary model sampler code to impute missing values in the pre- and post-test. Pupil-level KS1 average point score, everFSM, school-level average for KS1, and school-level average for KS2 were used as auxiliary variables. The results of the imputation are presented in Appendix E, Table 11 and Appendix F, Table 3. Despite imputation, the effect sizes remained consistent with the complete case analysis, although with minor changes to the intervals and variance. The treatment remained centred on 0 in the case of the primary analysis, and there was no change in the secondary outcome analysis where there was weak evidence of a positive effect.

Compliance and dosage

In this section, compliance is assessed against the criteria agreed with the developer: attendance at training, adequate staffing, initial assessment, termly review, pupil selection, and the delivery of support sessions in school (see Appendix H). In addition, the effects of dosage and composition of support are discussed.

Attendance at training, initial assessment, and termly review

Developer records showed attendance at training was very high with full compliance from 71 out of 75 schools (95%). Developer records also indicated that the initial assessment and termly review of pupils had been carried out in a large majority of schools with 68 out of 75 schools (91%, although data was missing for all but 1 of the remaining schools). Table 11 shows that roll-off across the two arms of the trial was broadly similar. The effects of roll off were investigated at pupil level by an interaction effect analysis using a three-category item, however, missing data reduced the sample to 1,076 pupils in 118 schools and none of the interaction effects were bounded away from 0 (autumn: -0.15, 0.47; spring: -0.23, 0.20; and summer: -0.41, 0.30). The results are presented in Appendix G, Tables 1 and 2.

Table 11: Descriptive statistics of the termly withdrawal (or roll-off) of numeracy support broken down by treatment group

Term that support was withdrawn	Catch Up® Numeracy	Active control
Autumn	105 (12%)	126 (14%)
Spring	160 (18%)	109 (12%)
Summer	456 (52%)	502 (56%)
Missing	155 (18%)	162 (18%)
Total	876 (100%)	899 (100%)

Dosage

Table 12 suggests that the length of numeracy sessions was more variable in the active control group, with a much greater proportion lasting more than 15 minutes, although active control schools may have provided the weekly numeracy support in one 30-minute session rather than two 15-minute sessions. However, comparison of data from the school logs for both groups suggests that the average number of sessions over the year was broadly the same in both groups (approximately 28 for Catch Up®; 33 for active control), giving an overall approximate session length in the active control group of around 23 minutes on average. However, these estimates should be treated with considerable caution since there is a substantial level of missing data, the data represents figures across the year (and does not account for different levels of roll-off), and the logs were collected in different ways in the two treatment groups. Unfortunately, the logs were not sufficiently similar to create a common robust scale across the two arms of the trial and the Catch Up logs did not provide information on the length of sessions. In addition, it was not clear whether all active control schools had followed the guidance on completing the logs and distinguished the numeracy support provided as part of the trial from other support for low attaining pupils.

Table 12: Length of numeracy support sessions in each arm of the trial, based on the TA survey

Treatment Group	How long do most numeracy sessions last?				
	Under 15 minutes	15 minutes	Over 15 minutes	Varies	Total
Catch Up®	0 (0%)	59 (54%)	30 (28%)	19 (17%)	109
Active Control	9 (8%)	11 (10%)	70 (64%)	19 (17%)	109

Composition (individual or group delivery)

From the data presented Table 13 below, it is clear that there was a difference in the proportion of group delivery between the two arms of the trial. In the active control, 49 out of 109 TAs (45%) delivered support sessions in groups most of the time, and a further 11 (10%) half of time, whereas in the Catch Up® intervention group delivery was almost entirely individual. The effects of composition were investigated at pupil level by an additional covariate using the five-category item from the TA survey. However, as the Catch Up® group were generally very consistent with their use of 15 minute sessions, the interaction effect analysis did not converge. The effect of the composition variable was not bounded away from 0 (-0.47, 0.16). The results are presented in Appendix G, Tables 3 and 4.

We carried out additional analyses on the entire dataset to investigate the impact of the use of group teaching versus one-to-one teaching as originally specified by the developer. We used a five category

indicator which asked teaching assistants retrospectively to report the use of one-to-one teaching across the year ('all of the time', 'most of the time', 'half and half', 'occasionally', and 'almost never'), with 'all the time' as our baseline category (see Appendix G, Tables 5 and 6). As above, the effects of the composition variables were not bounded away from 0. Hence, pupils experiencing group rather than individual delivery do not appear to have been disadvantaged.

Table 13: Descriptive statistics of the use of one-to-one sessions, from the TA survey

Treatment Group	Numeracy sessions are taught on the basis of one-to-one					
	All, or nearly all, of the time	Most of the time	Half and half	Occasionally	Almost never	Total
Catch Up®	107 (98%)	2 (2%)	0 (0%)	0 (0%)	0 (0%)	109
Active Control	34 (31%)	15 (14%)	11 (10%)	17 (16%)	32 (29%)	109

Pupil selection

Based on three measures collected—KS1 Average Points Score, Progress Test in Mathematics National Percentile Rank, and Overall Stanine—there were a considerable number of pupils selected outside the targeted attainment range. Table 2 in Appendix E provides quantiles for the above three measures. For the active control based on the assumption that eligible pupils should have scored below 15 on the KS1 average point score, 548 pupils were eligible (approximately 61% of those records matched), compared to 522 (60%) in the Catch Up® group. Hence, a large proportion (40%) of pupils selected were middle attaining or above at KS1. Indeed, some 19% of pupils with matched records had KS1 scores above expected levels and could, therefore, be classed as high attaining. Based on the PTM pre-test national percentile rank, only 45% of active control and 40% of Catch Up® pupils were within the expected attainment range of the bottom 15–20% of the attainment range. The distribution of prior attainment is discussed above (see Table 9) and presented graphically in Appendix E, Figures 2 and 3. At school level, only three schools in the control and six intervention schools selected pupils below age-expected level. In addition, 44 schools in the control and 45 schools in the intervention selected seven or more pupils under age-appropriate level, and lastly, 31 schools in the control and 28 schools in the intervention selected less than six pupils under age-appropriate level. While there is missing data at the KS1 level, this pattern is very similar with the PTM pre-test national percentile rank.

A per the protocol, analysis based on a comparison only of eligible pupils (that is, those with a KS1 score less than 15) found that the effect was not bounded away from 0 and remained consistent with the main findings, with an effect size of -0.08 (-0.27, 0.12). An interaction analysis indicated there was no relationship between group allocation and prior attainment. The results are presented in Appendix G, Tables 7 and 8.

The usual selection process involves a combination of professional judgment and assessments based on standardised tests and takes place after guidance is provided at the initial training. For this trial, this was not possible because pupil selection had to take place before randomisation and thus before training took place. Hence, whilst schools were provided with guidance by the developer, they were not provided with the normal training on how to implement this guidance. This may partially explain why a high proportion of middle attaining pupils were selected.

Summary

This analysis suggests that compliance to the Catch Up® Numeracy intervention was very high at school level for all criteria except pupil selection. This is a potential threat to the validity of the trial and is

discussed in the conclusion. It is worth noting again that, as highlighted earlier, misallocation by attainment is a very common (see, for example, Francis, Archer, Hodgen, Pepper, Taylor and Travers, 2017; MacIntyre and Ireson, 2002) and, hence, is likely to be an issue for any intervention where schools are required to select low attaining pupils. This issue was further investigated in the IPE (see Fidelity).

In addition, a large proportion of active control schools did not adhere to the guidance on 30 minutes of one-to-one support each week. However, the analyses conducted suggest that non-compliance on either the composition of support or pupil selection in either arm of the trial did not appear to affect the main findings. As discussed above, these results need to be treated with caution given the amount of missing data which has reduced the sample size significantly with these analyses. Unfortunately, the data on dosage was not sufficiently similar across the two arms of the trial to create a common measure that could be investigated statistically.

Cost

The cost estimate of £39 per pupil is based on the assumption of two TAs and one school coordinator attending initial training per school, as in the trial (see Table 15). Cumulative costs are presented in Table 16. These cost estimates do not take account of any additional training required where the original staff at a school leave and new staff require training. Additionally, Catch Up® provides additional review sessions at no additional cost to schools. The costs for travel to these additional sessions have not been estimated because attendance is not a requirement for the intervention.

Three half days' supply cover would normally be required for each staff member attending the training, including the school coordinator. Evidence from the survey and case studies indicated that, as was found on the previous evaluation (Rutt et al., 2014), additional time was required for planning and to access resources. Insufficient data was available to estimate any additional costs associated with this additional time. This issue is discussed in more detail in the implementation and process evaluation.

The previous evaluation (Rutt et al., 2014) indicated that some TAs reported that a large amount of photocopying was required. In the current evaluation, some TAs in the case study schools reported that, although initially they carried out a large amount of photocopying, this reduced considerably once the TAs became familiar with the resources.

Table 14: Cost of delivering Catch Up® Numeracy

Item	Type of cost	Cost	Total cost over 3 years	Total cost per pupil per year over 3 years
One-off initial training for TAs and school coordinator	Start up cost of £450 per trainee	£1,350	£1,350	
One-off travel cost to attend training	Start up cost per school	£50	£50	
Total		£1,400	£1,400	(£1,400/3/12) = £39

Table 15: Cumulative costs of Catch Up® Numeracy per pupil

	Year 1	Year 2	Year 3
Catch Up® Numeracy	£39	£39	£39

Implementation and process evaluation

In this section, the findings of the implementation and process evaluation (IPE) are reported, based on surveys and case studies. As already noted in the methodology section, the developer supplied data on support provided (including initial assessment) and on attendance at training. All other data collection and analysis was carried out by the evaluation team.

Respondents' and schools' identities are anonymised. The trial arms are referred to as Catch Up® (CU) and active control (AC), the respondents by their role (TA, teacher, school coordinator, or senior leader), and schools by a letter (A–E for each of CU and AC). The two TAs in each school are distinguished as TA1 and TA2. The survey results are presented in full in Appendix I.

In this section, the standard EEF headers are used because these are designed to address all the IPE dimensions and implementation factors highlighted in the EEF IPE guidance (Humphrey et al., 2016). The specific intervention-focused IPE research questions (RQ6-10) are addressed in this section as follows:

RQ6: To what extent do schools, coordinators, and teaching assistants perceive the Catch Up® Numeracy professional development for teaching assistants and coordinators to be effective? See sub-sections: *Were there any barriers to delivery being experienced?, Was the intervention attractive to stakeholders?*, and *What were the perceived outcomes of the intervention?*

RQ7: To what extent do the Catch Up® Numeracy intervention schools, teaching assistants, and coordinators adhere to the guidance and materials? See sub-section: *Fidelity.*

RQ8: Are children assessed for eligibility on a termly basis in both intervention and control schools and do children judged to have reached an age-appropriate level of numeracy 'roll off' the intervention? See sub-sections: *Fidelity,* and *Was the intervention attractive to stakeholders?*

RQ9: What does usual practice in control schools look like? See sub-section: *Control group activity.*

RQ10: To what extent does Catch Up® Numeracy have an effect on teaching assistants' attitudes towards mathematics and mathematics teaching when compared to an optimal business as usual active control? See sub-section: *What were the perceived outcomes of the intervention?*

Implementation

What were the necessary conditions for success of the intervention?

It is clear from the compliance analysis that—aside from pupil selection—fidelity to the intervention was largely high at a school level. In general, schools enabled staff to attend training and then implemented the key features of the intervention: initial assessments, termly reviews, and regular one-to-one support by TAs. The case study evidence suggests that school support and commitment for the intervention was a key factor in the successful delivery of the Catch Up® Numeracy intervention. These issues are discussed in some detail below. However, it is worth emphasising two aspects of the intervention that run counter to many schools' normal practice: off-site training for TAs and the delivery of TA support to individuals rather than to groups. Hence, school support in both of these areas appears to be particularly important. The extent of a school's support and commitment to TA training and individual support is likely to be strongly influenced by school priorities. The one case study school, School E (CU), that stopped delivery of Catch Up® during the year did so partially because, following an OFSTED inspection, the school decided that raising standards of literacy was a greater priority than numeracy, and partially because the school judged it more appropriate to organise support in groups rather than individually.

Were there any barriers to delivery being experienced?

Barriers that prevented the intervention being delivered as would have been hoped were identified by some survey respondents and case study interviewees. These perceived barriers are mostly those that would be expected given a new structured intervention delivered by TAs as they pertain to resources that are often in limited supply in schools—time and space.

TA workload and preparation time

The earlier evaluation of Catch Up® (Rutt et al., 2014) indicated that TAs could struggle to find the time they felt they needed to prepare. This problem was still encountered in this current trial. Indeed, the TA survey revealed that active control school TAs were more inclined to feel they had enough time to prepare for sessions (t = 2.2, df = 215, p = 0.028, d = 0.30)[10] although it should be remembered that Catch Up® schools were implementing a new way of delivering numeracy support whereas the active control schools were following an existing approach. Both groups felt equally strongly that preparation was necessary (t = 1.3, d = 212, p = 0.193, d = 0.18) and many TAs noted that the difficulty of finding quality planning time for their sessions was difficult irrespective of the actual intervention, summed up by TA2 in School D (CU): 'I think you always end up doing far more in your own time than you actually do in the working time because your working time is spent working with the children, and as much as you sort of need the time to prepare, you don't want that to impede on the quality time you spend with the children.' A TA in School C (CU) noted this was less problematic if the pupils had similar needs: 'It will get bigger, depending on different children, I was quite lucky that the six children that I got have got the same gaps missing.' At least one case study school responded by reducing the number of pupils that TAs saw back-to-back or reorganising the timetable to allow planning time at the beginning of sessions.

Interviews with staff in the case-study schools suggested that both initially finding and then understanding Catch Up® resources required preparation time. For example, 'Yes when I first started I kept going on the website and kept looking and printing bits off but that is as far as it went because then I just didn't have the time to sit down and look at it, I did start taking it home' [TA1 School E (CU)] and TA1 from School A (CU) suggested: 'But the only problem we have had with them—and this is I think a problem for everybody—is because they are online, you have got to have time to find them, print them out, prepare them.' Furthermore, five of the eight interviewed TAs and one coordinator mentioned that resources could be complicated to understand. TA2 from School D (CU) commented, 'I must admit I don't actually understand, because you sort of read it [lesson plan] and you think what are they actually asking you to do.' This contrasted with the usual practice in School D (CU) as the TAs typically had face-to-face discussions concerning what to do with a teacher rather than reading through lesson plans on their own.

Difficulties identifying a suitable space outside the classroom

The earlier evaluation of Catch Up® (Rutt et al., 2014) noted the difficulty of finding a suitable location within the school to run the intervention. This remained a challenge in this trial. The TA survey showed that 57% of the Catch Up® schools reported that most sessions were held in a separate classroom and all were held outside the standard classroom. It should be noted that 17% of control school TAs reported that they held sessions within the standard classroom. Interviewees commented that finding a quiet space could be challenging and TA2 from School E (CU) reported that this situation could be exacerbated at certain times of the year: 'There are very few places, especially coming up to SATS because Year 6 seem to take all of the spare space, so it is normally done in the corridor which as you can imagine with people coming and going.' Even when they did have a quiet location, the practical barrier of time spent fetching individual pupils to it was reported as an issue by a number of TAs; for example, 'Yes when you have got five and I can't always do it in this room so sometimes I have to go

[10] Welch's test for unequal variances is used. Degrees of freedom (df) are rounded to the nearest whole number.

to the far end of the school and just walking up and down that corridor that takes time so then you're losing time then off' [TA2 School B (CU)].

Was the intervention attractive to stakeholders?

One key indicator of the intervention attractiveness to stakeholders is whether they would like to continue with it the following year. Respondents to the TA, coordinator, and senior leader surveys in both Catch Up® and control groups were positive about their experiences and most indicated they wanted to continue next year. For example, only 14% of TAs, 8% of coordinators, and 6% of senior leaders expressed a negative view about continuing with Catch Up® the following year. However, there was only a single significant difference between the conditions and this was the coordinators from Catch Up® schools were keener to continue next year than their equivalent in the active control schools ($t = 2.4$, $df = 121$, $p = 0.019$, $d = 0.42$).

There were a number of aspects of Catch Up® that were identified as attractive. When responding to the question, 'What are the perceived outcomes of the intervention?', one reason for continuing—perhaps the key reason—was the perceived benefits for pupils including numeracy attainment and confidence. Beyond that, though, the most frequently discussed reasons for finding Catch Up® numeracy sessions attractive concerned the themes, the structure of sessions, and personalised learning.

The structure of support sessions

Although both Catch Up® and control schools reported using a common structure across the sessions, the TA survey indicated that Catch Up® increased this tendency ($t = 3.0$, $df = 190$, $p = 0.003$, $d = 0.41$). Consequently, it is positive that many of the staff interviewed were very appreciative of the Catch Up® structure; for example, TA2 from School A (CU) noted that her favourite thing about the intervention was 'the structure, the actual structure, the way that it's structured so well'. TA2 School C (CU) suggested that pupils value it too, 'because I think some children find it easier with having that set structure, they know how things are going to work so there is no mysteriousness about it, they know they are going to come in and how the lesson is going to go'. The coordinator at School A (CU) considered the 'short, sharp, and focused' structure of the intervention was the reason it was beneficial. Some felt this tight structure could support a TA's lack of knowledge: 'The staff didn't have the knowledge, the background knowledge or their own confidence; it is quite nice to have the referral point to go to' [Coordinator School B (CU)].

Personalised learning

The other favourable aspect of Catch Up® identified by some TAs, teachers, and coordinators was it ability to personalise the learning to a pupil's specific needs by focusing on the gaps in their underlying numeracy. For example, the coordinator from School C (CU) thought Catch Up® was an enhancement to the school's previous approach 'because it is more focused and it is actually plugging gaps that they have got in their understanding'. This view was echoed by teachers, for example, teacher School A (CU): 'What they are doing in the Catch Up® is completely different to what we're doing in the lesson; it [provides] those building blocks of maths that the children need.' TAs felt they therefore offered a valued complementary activity 'The advantages are [that it identifies] those gaps there, and whilst [the pupils] probably can add two fractions together, they can't count to 50 or back from 50 in 2's, so I kind of think it fits what it is meant to do—it is filling in the gaps that they haven't got.' This was seen as a distinctive attribute of Catch Up® compared to other interventions, by, amongst others, TA2 in School C (CU): 'I think Catch Up® is quite on its own compared with other interventions and […] is totally different in that it is filling gaps rather than going over what they have done in a lesson.'

A structured numeracy support programme

Compared to the Catch Up® schools, senior leaders (t = 2.5, df = 93, p = 0.016, d = 0.50) from the active control schools had more of a tendency to perceive that a highly structured intervention would be easier to deliver. This was not the case for the coordinators or the TAs from the two types of schools, as there were no differences in their perceptions (t= 0.5, df = 122, p = 0.642, d = 0.08; t = 0.3, df = 213, p = 0.773, d = 0.04).

There were also some features of Catch Up® that were less attractive to stakeholders.

Individualised support

Of key concern was the requirement of one-to-one support rather than small group support. Although all case study schools did implement numeracy support as one-to-one, three of them stated that this was a reluctant departure from their usual practice, and the following year, School B (CU) and School C (CU) intended to change to run Catch Up® in small groups. School E (CU) did not intend to run Catch Up® at all. A desire to move away from a one to one support model contributed to this decision. Although there were multiple, interdependent reasons for the preference for a group intervention, the dominating one was the costs of reallocating TA time to both deliver and prepare the sessions. This concern was succinctly summed up by the coordinator from School B (CU): 'The timing of it … I think it is very restrictive for us, so I think probably we will run it, but more as a group—put them in to a group to get more children.' Two other reasons that were given for preferring small group were the advantages of pupils 'feeding off each other', and a concern that one-to-one can be daunting for some children, particularly when they are not familiar with the TA. A coordinator from School B (CU) suggested that, depending on the needs of the pupils, there could be flexibility in the composition of a group. One week, the TA may deliver to just one pupil, but the next week, if several children have encountered a similar issue, they may all be delivered the intervention.

Time taken for the initial assessment

The amount of time taken for assessment in the first term of the intervention could be considerable. The Catch Up® Numeracy guidance (Catch Up®, 2017) estimates that the initial assessment should take between 60 and 90 minutes and suggests that this can be carried out in one longer session. However, the assessment was generally administered within the twice-weekly 15-minute sessions and so, as a result, in some schools, and for some pupils, the assessment continued beyond the first half term break, and may in some cases have contributed to the assessment taking longer than expected. Interviewees in all the case study schools expressed surprise and a little frustration that the assessment took longer than expected for some pupils. They also were concerned about the impact upon pupils who could find it 'daunting because it was page after page after page [of assessment]' [TA1 School E (CU)]. In fact, the records supplied by the developer indicate that the mean time taken for the assessment was approximately 80 minutes and, for the majority of pupils (72%), the assessment was completed under 90 minutes. However, the assessment did take 150 minutes or longer for some pupils (8% of all pupils involving 29% of CU schools). It is noted that the developer considers that the assessment contributes to pupil support and, thus, should not be regarded simply as assessment. This, as discussed below, could be made clearer by Catch Up® in their initial training.

Fidelity

Was the intervention delivered as intended to all in the treatment group? If there were any issues with fidelity, what were the reasons?

In this section, the extent the intervention was delivered as intended is discussed, together with the reasons for deviations. As well as survey and case study findings, results from the impact evaluation

are drawn on in order to explore how the key features of Catch Up® were implemented. This section should be read in conjunction with the earlier compliance analysis.

Training

Attendance at the initial pre-project training sessions was very high at 97%. However, when surveyed at the end of the year, a slightly low proportion of TAs and coordinators reported that they had attended the initial training: 87% and 84%, respectively. This is probably due to a number of TAs and coordinators being replaced during the year. Catch Up® review sessions also occurred throughout the year; 67% of surveyed TAs and 66% of surveyed coordinators reported attending.

Dosage

Catch Up® is designed to run consistently in two 15-minute sessions per pupil per week. Most respondents in Catch Up® schools stated that efforts were made to ensure each child did receive the required dosage. Indeed, the coordinator survey showed that 92% of the Catch Up® coordinators and 95% of control school coordinators reported the intervention sessions were scheduled into the pupils' timetables. This situation was largely reflected in the data derived from Catch Up® case study schools. However, as already discussed in the compliance section, evidence on dosage from the active control schools suggests that dosage in those schools may have been higher than in the intervention schools.

As was reported in the earlier evaluation of Catch Up® Numeracy (Rutt et al., 2014), school priorities external to pupil's progress in numeracy appeared to impact on the frequency of sessions. These influences generally entailed a decrease rather than an increase in dosage. Seasonal factors such as Christmas activities and school trips led to sessions being missed as pupils were elsewhere. Other interventions could also compete for pupil's time. Furthermore, TAs may be unavailable to deliver sessions. For example, in case study schools, some more qualified teaching assistants missed sessions when they were asked to cover lessons for a teacher.

The extent that efforts were made to replace the missed sessions appeared also to be determined by the same combination of interrelated circumstances with the cost of reallocating the TAs' time emerging as a central factor, fuelled possibly by the extent the senior management were committed to the intervention. For example, a supportive deputy head who adopted the role of coordinator in School A (CU) commented: 'I have said to the TAs we are going to commit. We are not going to miss sessions.' In all case study schools, however, it was largely the TA rather than coordinator who was responsible for ensuring the weekly dosage was consistent.

Pupil selection

For this project, the pupils eligible for the intervention were those that were considered to be amongst the lowest attaining 15–20% of children for their age. The compliance analysis indicated that only around 60% of pupils selected were within this subgroup of pupils. As discussed earlier, one reason for this may be that, in this trial, in order that pupil selection was blind to group allocation, the selection of pupils took place prior to the normal training on the initial selection of pupils. The case study data suggests additional reasons why some low attaining pupils may not have been selected. Two reported that they adapted the criteria to suit their own contextual needs. For example, the coordinator in School A was reluctant to include any pupils who had recently been involved in an intervention—'they needed a break'. Although her criteria did include pupils' attainment in numeracy, she was also concerned to include those pupils 'lacking confidence, or looking unhappy or unsure about maths, or just not engaged'. The coordinator from school B cited behavioural issues were a reason for a couple of pupils attending the Catch Up® sessions. Some respondents appeared to recognise some issues with selection. For example, a class teacher from School E (CU) explained she did not think the intervention was appropriate for all the pupils as 'some of them, I think, were quite capable to start with'.

Scheduling of support sessions

Scheduling was approached similarly in the active control schools to Catch Up® school: 95% of active control schools reported scheduling support sessions into the pupils' timetables (compared to 92% of Catch Up® schools).

'Rolling off' and termly review

Ensuring pupils were rolled off the intervention when they achieved the numeracy skills appropriate to their age was recommended to the schools participating in the project. The earlier compliance analysis indicated that compliance to this requirement was high. In line with this, the coordinator and senior leader surveys revealed no difference between how coordinators from the control and Catch Up® schools perceived the criteria to roll pupils off the intervention: 8% and 6% of senior leaders and 16% and 16% of coordinators, from the Catch Up® and control schools respectively, reported that the decision to end the intervention for a child was based solely on a standardised test. In contrast, 56% and 57% of senior leaders and 59% and 41% of coordinators, from the Catch Up® and control schools respectively, reported that a combination of professional judgement, standardised test, and staff resourcing were all part of the criteria used to decide to roll pupils off the intervention.

Individual delivery of support

Catch Up® is designed to be an individualised programme as compared to working with a group of pupils; this composition may reduce the pedagogic demand for TAs, and hence increase the likelihood of the sessions being effectively delivered. The TA survey indicated that TAs from the Catch Up® schools reported that 98% of their sessions were delivered one-to-one, which stands in stark contrast to control schools where group session were much more frequent ($t = 12.3$, $df = 216$, $p < 0.001$, $d = 1.67$). This result was echoed in the evidence derived from the case study schools. During the period of the intervention, all the Catch Up® case study schools implemented numeracy support as one-to-one.

What elements of the intervention were perceived to be adaptable?

As discussed above, the Catch Up® schools followed the advice from Catch Up® closely during the period of the intervention and appeared to do so with little desire to adapt most aspects of the intervention. However, as discussed in Question 3—'Was the intervention attractive to stakeholders?'—this was not true of the Catch Up® requirement of one-to-one support as two case study schools reported intending to move to group delivery.

In addition, some case study TAs suggested minor adaptations or 'tweaks' to the timing, structure, and length of the support sessions. These suggestions appeared to be in line with the developer's views.

Outcomes

What were the perceived outcomes of the intervention?

The perceived outcomes of the intervention were analysed for both pupils and staff. Clearly, a key learning outcome of the intervention was to develop pupils' numeracy skills. However, as the extent to which pupils are confident in numeracy can influence their learning of it, this too was a focus of the analysis. When considering staff outcomes, it was judged important to investigate whether the experience of delivering Catch Up® influenced TA professional development. Additionally, how the intervention impacted on communication between staff was explored. These four factors were also the subject of investigation in the previous evaluation of Catch Up® (Rutt et al., 2014), and are detailed, in turn, below.

Pupil attainment in numeracy

Overall, the TA, coordinator, and senior leader surveys showed that both Catch Up® and active control schools considered that their pupils had improved at numeracy during the intervention. However, there were no detectable differences between the two conditions according to the TA, coordinator and senior leader surveys (t = 1.7, df = 210, p = 0.087, d = 0.23; t = 0.3, df = 113, p = 0.767, d = 0.05; t = 0.4, df = 95, p = 0.696, d = 0.08).

The case study school interviews were able to look more closely at what attainment may mean by exploring perceptions of progress through the levels of the intervention and advances in classwork. There was generally a consensus that pupils had progressed through the intervention. However, views were more diverse regarding progress in class, with the coordinator from School E (CU) stating: 'They [teachers] don't feel that it is having too much of an impact on the children's actual learning within the class.' Other participants reported that because some pupils' understanding of basic numeracy was fragile, the intervention was unlikely to have immediate impact on their class work. Pupils may make substantial progress in Catch Up, but not quickly attain age-appropriate numeracy skills. This perception was captured by the coordinator from School A (CU) who commented: 'The children are playing Catch Up®, for want of a better phrase, you know they are not there yet, they are not … they have not caught up to Year 5 yet, you know there still is that gap between them and their peers and I would say probably the biggest difference I have seen has been in this last bit of the year.' Staff from Schools B (CU) and E (CU) considered that the intervention would have a bigger impact on pupils' classwork if the children were drawn from lower years.

When pupils did make progress in their everyday lessons, participants, to varying degrees, believed, at least in part, this was due to Catch Up. A TA from School A (CU) supposed the two-year progress a child had made within six months was because of the impact of the intervention. Participants also considered that other factors could influence progress. Some pupils, for example, received extra attention in class, and in three schools some participated in another mathematics intervention. A teacher from School E (CU) comments: 'I don't think it [pupil progress] was necessarily due to Catch Up® maths because they are making progress in fractions … which isn't what they have been covering [in the intervention].'

Pupil confidence in numeracy

The TA, coordinator, and senior leader surveys concerning pupil attitude to maths showed a similar pattern as above. These all reported a strong improvement in student attitude to numeracy. There were no differences between the control and intervention schools, although a tendency to enhanced engagement was reported in the coordinator survey (t = 1.9, df = 116, p = 0.063, d = 0.34). Teachers, TAs, and coordinators from all the case study schools focused on pupils' growth in confidence as a key outcome of the intervention. Confidence was mainly considered in terms of numeracy specifically, although some participants also mentioned general growth in confidence. Evidence of pupils' confidence in their own mathematical capabilities was drawn from a number of behaviours. This included pupils answering questions in class, being more resilient when faced with a challenging problem, and asking for help when stuck.

Teaching assistants' attitudes to mathematics, mathematics teaching, and professional development

It was hoped that working within Catch Up® supported by training and structured resources would have a positive impact upon the TAs' professional development in comparison to the active control. However, although satisfaction with the training was very high, the results from the TA survey suggest that the TAs in the Catch Up® group did not make these anticipated gains. TAs in the intervention group reported less enjoyment of teaching maths (t = 2.3, df = 203, p = 0.024, d= 0.31), less enjoyment of maths (t =

2.1, df = 210, p = 0.035, d = 0.29), less improvement in teaching maths (t = 2.6, df = 210, p = 0.009, d = 0.36), and less improvement in their mathematical knowledge (t = 2.4, df =214, p = 0.016, d = 0.33). The senior leader and coordinator surveys reported no differences in their perception of TA professional development. It should be noted, though, that these results do not mean that TAs are perceived to not have developed. The surveyed respondents in Catch Up® and active control schools generally were positive about the opportunities provided in both contexts for such development.

When considering training—as discussed in, 'Was the intervention delivered as intended?'—attendance was very high. Moreover, this was seen as a distinct benefit as there was general consensus that typically TAs lacked opportunities for professional development and rarely attended training events.

Moreover, a general positive consensus could be seen concerning this training. The vast majority of responses concerning the trainer's knowledge, clarity, and the usefulness of materials were rated very positively (for example, over 98% of responders indicated they found these 'excellent' or 'good'). The least positive perceived aspects of the training were simply practical issues concerning the venue and hospitality (scoring a lower 76% positive rating). The vast majority (over 94%) of attendees were also positive about their learning from it, feeling either 'strongly confident' or 'confident' they could put Catch Up® into practice themselves. Looking at their free text responses, attendees most liked the videos of best practice, explanations of the structure of the intervention, and how to deliver a lesson. In terms of improvements, the most frequently mentioned aspects were a desire to change the training from three half days to one full day, less time on background research, and more on practical activities, although not all agreed with this. For example, the coordinator from School D (CU) did have some concern that Catch Up® training emphasised the processes TAs needed to attend to in order to implement the intervention, rather than the underlying mathematical difficulties pupils face: 'There is no support actually on teaching maths or the learning of maths but there is support on how to manage an intervention.' Catch Up® review sessions occurred throughout the year; 67% of surveyed TAs and 66% of surveyed coordinators reported attending. Again, these sessions were generally described in positive terms. Participants found they provided a reassuring opportunity to check whether they were doing the 'right thing' and, in particular, to share experiences and common difficulties such as timetabling.

An indirect outcome of being involved in the intervention was, for some, improved job satisfaction. A TA from School D (CU) commented: 'It's very rewarding for me to see them improve and for them to get something, even the smallest step, it does feel a real sense of achievement.'

Communication between staff

Typically, research suggests that in schools there is often only a limited amount of communication between teachers and TAs (Sharples et al., 2015). This section considers both the frequency and the nature of communication in the Catch Up® schools. There was a general recognition of the importance of communication. This perspective was captured in the comments by a senior leader from School D (CU) that 'communication is huge because if you don't share that knowledge and those experiences and the things that have worked and the things that haven't worked, you're missing the trick'. Unfortunately, the intervention did not appear to be a catalyst for prompting more communication between TAs and class teachers. Indeed, the TA survey results suggest that Catch Up® TAs talked less to teachers about the numeracy sessions as only 45% talked to teachers weekly compared to 78% of their control group equivalents. Face-to-face dialogue was generally ad hoc and 'in passing'. Scarcity of time was the overriding reason for this situation. Underlying there appeared to be two interrelated factors. First, when TAs did not physically work in the teacher's class there was limited communication between teachers and TAs. As one TA commented, in this situation, they were more likely to make just the 'odd comment' to the teacher about a child's progress. Second, many participants perceived there was no need to connect the learning in the intervention to pupils' classwork. As TA1 from School C (CU) asserted, 'Rather than going by what they are doing in class, I am doing what gaps they are missing.' Thus, the incentive to discuss their learning was reduced. Indeed, none of the TAs in the case study schools prepared work for the sessions with the teacher; they

decided for themselves what to do. The perception that Catch Up® is self-contained, separate from pupils' everyday work was summed up by a TA from School A(CU): 'Not even the teachers are really familiar with it [Catch Up®]; I mean, I once had a teacher ask me, "Oh, when you take ... out for his Catch-Up, can you work a little bit on something?", and I said, "Well, no, because that's not how it works".' However, in School D (CU) this problem was not as acute because the coordinator arranged meetings between herself, teachers, and TAs to tell them about the gaps in pupils learning that Catch Up® had uncovered. These conversations were not only initiated by the coordinator: teachers in this school reported approaching the coordinator to suggest the focus for an intervention session.

Communication between TAs and coordinators seemed to be more frequent. There was no strong difference between the intervention and control schools according to the coordinator survey (which reported weekly conversation at 48% in Catch Up® schools and 39% in control) and TA surveys (35% and 31% respectively). A further means of communication was through observations. Two coordinators and one senior lead had observed sessions. There is some evidence from the senior leader survey to suggest that in Catch Up® schools sessions were observed slightly more frequently: 74% of senior leaders reported such observation compared to 69% in active control schools.

Were there any unintended consequences or negative effects?

There could be both unintended and negative consequences for staff and pupils because of their implementation of Catch Up, therefore both these themes were inspected for possible effects.

In terms of pupil consequences, the earlier evaluation of Catch Up® (Rutt, Easton and Stacey 2014) had noted negative attitudes of pupils to the intervention with some pupils reporting feeling stigmatised by their inclusion. However, this did not seem to be an issue in the current trial as none of the case study schools reported this effect, even when specifically queried. For example, the coordinator from School D (CU) responded, 'Children come out to do additional activities so just because you're being taken out of class doesn't label you, we take them all.' The only negative pupil attitudes related by interviewees were from School E (CU) (the school that stopped Catch Up) who commented that some pupils found the intervention dull 'when they get up to Year 5 because I think a couple of the children I think to start with were "oh this is boring"' [TA1, School E(CU)].

In terms of unintended consequences for staff—as discussed in the question (Q2), 'Are there any barriers to delivery being experienced?'—TAs in Catch Up® schools reported increased workload due to the time needed to prepare for the sessions. In addition, this had negative consequences beyond the TAs themselves. A teacher in School E (CU) who missed the support they received commented: 'So if they were having planning time somewhere and I find it really useful to have my TA actually in my maths lessons to support, especially having set 2, the more adults you can get [in the lesson] the better.' This may have contributed to their decision to withdraw. They were not alone in this concern. The TA survey revealed that TAs in the intervention group felt that their time out of the classroom had a much greater negative impact upon classroom management ($t = 3.9$, $df = 209$, $p < 0.001$, $d = 0.53$), although this was not echoed strongly in the coordinator and senior leader surveys.

Formative findings

Are there any ways that the intervention can be improved?

Although the majority of feedback was positive (for example, see the high numbers wishing to continue next year), participants did refer to a few specific aspects of the intervention that could be improved.

As explored through the question (Q3), 'Is the intervention attractive to stakeholders?', many TAs were surprised and concerned about how long the initial assessment took. The recalibration of this expectation to a more realistic one can easily be achieved. The initial training sessions provided by

Catch Up®, and the Catch Up® Numeracy booklet, could highlight that the initial assessment may take longer than 90 minutes, particular when TAs are new to the programme.

There was mixed enthusiasm expressed for the resources. Some made very positive comments, however, there were concerns expressed around the practical issues of access and media. Several TAs expressed difficulties accessing the online resources. They found navigating the website time consuming. TA1 from School D (CU) commented: 'I think the titles are too generic really […] you've got to open each file individually to see whether that one is actually suitable for you.' The website could be altered to minimize the need to open files in order to understand broadly, their content, or providing an HTML version.

These findings are similar to those reported in the earlier evaluation (Rutt et al., 2014), which recommended improvements to the initial assessment and the provision and style of the resources. However, this study did not appear to find the same degree of concern for the number of resources that Catch Up® provides or their 'modernity' suggesting this is not perceived to be an issue.

The active control schools were advised to use the EEF's TA Guidance Report (Sharples et al., 2015) and follow its recommendations to inform the support provided to pupils. In other respects, the support provided by the active control schools was similar to the Catch Up® intervention. The pupils selected were deemed to be struggling with numeracy prior to randomisation (as judged by the schools). The recommended dosage was 30 minutes per week (either one 30-minute session or two 15-minute sessions) in one-to-one sessions, with pupils rolling off the intervention when they had attained the age-appropriate level in numeracy (determined by the school). Each participating school had the freedom to decide for itself the structure and content of the sessions. However, the schools' choices may have been partly influenced by the research in that the evaluation team had regular contact with active control schools to collect logs of the number and timing of support sessions.

Control group activity

What happened in the control group and how did this compare to what was intended? (In particular, was there any resentful demoralisation or compensation rivalry going on? What was 'business as usual'?)

The following section uses the data from the TA, coordinator, and senior leader surveys to describe overall patterns, with the five control case study schools providing opportunities to probe these patterns more deeply. In reporting here, to keep the length of this section manageable, only those aspects of the intervention where control schools differed markedly to intervention schools are discussed.

Dosage and composition

The most significant points to note relate to group composition and dosage in the active control group, where support sessions were more frequently delivered in groups and overall dosage may have been greater. These issues have been extensively discussed elsewhere in this report (see the section on compliance and elsewhere in the implementation and process evaluation).

Assessment

The proportion of pupils for whom coordinators reported having completed an initial assessment in the Catch Up® and control schools was very high and similar (92% and 86%, respectively, reported that all pupils had completed an initial assessment). However, results from the TA survey found TAs in the Catch Up® group felt the initial assessment had greater influence on the content of the sessions than did the control group TAs ($t = 7.3$, $df = 159$, $p < 0.001$, $d = 1.00$). Correspondingly, the TAs from the control group considered the class teacher's assessment of pupil learning was a greater influence on the content of the sessions ($t = 4.6$, $df = 214$, $p < 0.001$, $d = 0.62$).

Connection to everyday lessons

The case study Catch Up® schools' data suggested that the numeracy intervention sessions ran autonomously alongside, but not integrating with, regular classroom teaching. However, the TA survey indicated that the content of resources used in active control schools was more often the same as those used in class ($t = 5.1$, $df = 214$, $p < 0.001$, $d = 0.69$).

Resources

Compared to the Catch Up® schools, control school respondents to the TA survey felt the resources used in their intervention sessions were more engaging ($t = 6.8$, $df = 192$, $p < 0.001$, $d = 0.92$). They also reported using more resources that were the same as those used in a pupil's recent maths lessons ($t = 5.1$, $df = 214$, $p < 0.001$, $d = 0.69$). These results may, in part, be due to the fact that, as discussed above, class teachers had a greater influence on what happened in the sessions delivered in the control schools.

Training

The TA survey indicated that control school TAs had not had as much training, with nearly a third reporting no training at all. In contrast, the coordinator and the senior leader surveys showed no difference between control and Catch Up® schools and described range of training they had undertaken. The support the EEF provides regarding how to effectively deploy TAs within schools was utilised to varying degrees.

Attitude to being part of the control group

A final issue to briefly consider is the how staff in the active control schools felt about their allocation to this condition. Members of three of the five schools expressed disappointed at being assigned to the control group. The coordinator from School A (AC) commented that she did not want to be in the control group because they wanted to learn about another approach to supporting pupils struggling with numeracy. For School B (AC), being assigned to the control group had appeared to prompt them to purchase a standard structured intervention programme. The coordinator of School D (AC) expressed they had hoped to be part of the Catch Up® group because they did not have a structured intervention programme for numeracy and 'we wanted to improve how we were using our TA's within the classroom, and to give them more training really'. To compensate for this, she spent time with the TAs creating plans for the one-to-one sessions. However, this ended after a few months when it was perceived not to be effective.

Conclusion

Key conclusions	
1.	There is no evidence that Catch Up® Numeracy had an impact on pupils' maths outcomes, on average. This result has a low to moderate security rating.
2.	Among pupils eligible for Free School Meals, those who received Catch Up® Numeracy made two months less progress in maths compared to those who did not. This result may have lower security than the overall finding because of the smaller number of pupils.
3.	There was some weak evidence to indicate that children who received Catch Up® Numeracy developed a more positive attitude towards maths compared to those in the active control group.
4.	Schools implemented the intervention with a high level of fidelity. However, there was some mis-selection of pupils which suggests that intervention programmes for low attaining pupils may not reach all pupils for whom they are intended.
5.	Teaching assistants reported that they valued the Catch Up® Numeracy training. However, there was no evidence that Catch Up® Numeracy changed the attitudes of teaching assistants to maths and maths teaching.

Interpretation

The findings of the impact evaluation suggest that there is no evidence of the impact of the Catch Up® Numeracy intervention on pupil attainment compared to the optimal active control. This was in line with the findings of the previous trial (Rutt et al., 2014) in which Catch Up® Numeracy was compared to a matched time control, and similarly did not find an effect. This strongly suggests that the potential cross-contamination effects were not a practically significant factor in the previous result. However, unlike the previous trial, it was not possible to estimate the effects of the intervention when compared to a 'no support' passive control. Considered together, the results of this current trial and the previous trial (Rutt et al., 2014) suggest that Catch Up® Numeracy is no more effective than an active control condition involving regular support by TAs, but the previous trial suggests that both appear to be more effective than a no support condition.

Additionally, there was evidence to suggest that the intervention did not impact negatively on the attitudes or motivation of some pupils as was highlighted in the previous evaluation (Rutt et al., 2014), and, indeed, there was some weak evidence to suggest that the intervention may have had some positive impact on pupil attitudes.

The implementation and process evidence indicates that, aside from pupil selection (discussed under limitations below), fidelity to the intervention was high at school level for Catch Up® schools; schools delivered regular support to pupils through TAs, enabled their staff to attend the required training, and adhered to termly assessment and review. The IPE evidence also suggests that active control schools provided regular ongoing support in line with the requirements. As discussed below, there was less compliance with the dosage and the composition of support requirements amongst the control schools. It is worth noting that taking part in the trial may have been a factor in the provision of this ongoing support since schools were contacted by the evaluation team on a half-termly basis to ask for records of the support provided to pupils.

The cost evaluation indicates that Catch Up® Numeracy is a low cost intervention. This estimate (£39 over three years) differs from the cost estimate in the previous evaluation (Rutt et al., 2014) because it is based on the direct costs, marginal costs, and pre-requisite costs; the costs of TA and school coordinators' salaries are not now included as per the current EEF guidance on cost evaluation (EEF, 2016). This evaluation found no evidence that additional resources were necessary. The case study

data suggested that some additional photocopying was required, although the costs of this appeared to be minimal. Evidence from the survey and case studies indicated that, as was found in the previous evaluation (Rutt et al., 2014), in comparison to the active control, additional time was required for planning and to access Catch Up® Numeracy resources.

The survey results indicate that both 'interventions', Catch Up® Numeracy and the active control support, were attractive to schools, and staff in both arms reported improvements in pupil attainment and attitudes. In addition, there was no evidence that participation in the Catch Up® Numeracy intervention led to an impact on the attitudes of teaching assistants to mathematics and mathematics teaching when compared to the active control, although teaching assistants appeared to value the Catch Up® Numeracy training provided. Survey results indicated that implementing Catch Up® Numeracy resulted in less communication between TAs and class teachers compared to the active control group.

The earlier evaluation of Catch Up® (Rutt et al., 2014) noted the difficulty of finding a suitable location within the school to deliver support sessions and that school priorities external to pupils' progress in numeracy appeared to impact negatively on the frequency of sessions. The case study evidence suggested that both issues remained a challenge in this trial.

The earlier evaluation (Rutt et al., 2014) recommended improvements to the style of the resources and the organisation of the resources on the Catch Up® Numeracy website. Case study evidence suggested that accessing resources on the website continued to be perceived as a problem. However, the current study did not appear to find the same degree of concern about the style and content of the resources.

The implementation and process evaluation suggests that many schools have a strong preference for delivering numeracy support in groups rather than one-to-one sessions, and that many active control schools delivered support in groups. Survey data indicated that the majority of schools in the Catch Up® Numeracy group intended to continue delivering the support after the end of the trial, although evidence from case study schools suggested that many schools may deliver support in groups rather than one-to one.

Limitations

The first limitation relates to activity in the active control group. Although compliance was high in the Catch Up® Numeracy intervention group, the process evaluation data suggests that, in the active control group, support sessions were more frequently delivered in groups and overall dosage may have been greater. Increased group, rather than individualised, delivery of support could result in a differential impact on attainment, and, indeed, somewhat counter-intuitively, Pellegrini et al.'s (2018) meta-analysis indicates that small group delivery of tutoring programmes has a slight advantage over one-to-one delivery.[11] However, the inclusion of an individual vs. group delivered categorical covariate showed a null effect. Increased dosage could also result in a differential impact, although, as described earlier, it was not possible to investigate this quantitatively because dosage data was collected differently between the two treatment groups and, despite considerable efforts from both the evaluator and developer teams, there was a substantial amount of missing data. Future trials could consider sampling dosage for particular time periods rather than collecting comprehensive dosage logs, which might result in more complete and valid data as well as being more cost effective.

A second limitation relates to pupil selection. Over 50% of the pupils selected for the trial were within the bottom 15–20% of the attainment range. The IPE indicated that in some cases eligible pupils were identified but not selected for the intervention due to concerns about intervention fatigue, pupil confidence, or behavioural issues. Such mis-selection is not surprising given the literature on allocation

[11] At the time of writing, Pellegrini et al.'s (2018) meta-analysis and best evidence synthesis was only available as a summary conference paper. See also Slavin's (2018) discussion.

to groups by attainment (see, for example, Francis et al., 2017). This mis-selection represents a potential threat to validity since the intervention was not intended to help pupils who were not low attaining. However, a per-protocol analysis at pupil level indicated that this was not a serious issue because the result was consistent with the main findings.

A third limitation relates to attrition. Although attrition at school level was low (5%), attrition at pupil level was relatively high (17%). This was largely due to two factors: first, a large proportion of pupils moved schools at the end of the 2016/2017 academic year after the trial ended but prior to the administration of the delayed post-test, and, second, a significant proportion of the pre- and post-tests collected and marked by GL Assessment were missing and not matched to pupils. This level of attrition raises some uncertainty about the security of the findings. However, the sensitivity analyses suggest that the results were not affected by the attrition. Attrition is likely to be an issue for similar delayed post-test designs. For future trials, it is recommended that evaluators take this into account this when considering research designs involving delayed outcomes.

A fourth limitation relates to imbalance within the sample, the limitations of the GL PTM tests, and pupil selection. The trial involved pupils across three years groups (aged 6–10) and, as a result, it was necessary to assess the impact using tests appropriate across this age range. Although the test provider, GL Assessment, advised that the relevant PTM tests, and age-adjusted scores, were appropriate for pupils at all attainment levels including the low attaining students targeted by Catch Up® Numeracy, a very significant floor effect was evident. This may have introduced some bias by diluting the detectable effects of the intervention. This was investigated by fitting a Bayesian equivalent of a Tobit model which adjusted for the floor effects. While this produced similar pupil-level effects, along with an increase in the observed variance, the floor effects remain a significant loss of information. In addition, a large proportion of the pupils selected in both arms of the trial were middle and high attaining pupils rather than pupils struggling with numeracy. This may have introduced some bias in that the active control support may have been more appropriate for these pupils. A per protocol analysis was carried out and the impact of the treatment remained a null effect. Nevertheless, although no evidence of bias was found, the level of imbalance raises some uncertainty about the security of the findings.

A fifth limitation relates to the design of the trial. In this trial, the Catch Up® Numeracy intervention was compared to an active control. A three-arm trial was originally proposed with an additional ‘no support’ passive control. Following discussion with the developer, this option was judged not feasible due to the large number of schools that would be required to adequately power the trial. As a result it is not possible to robustly estimate the effects of Catch Up® Numeracy compared to a ‘no support’ option, although there is evidence of an effect from the previous trial (Rutt et al., 2014).

A sixth limitation relates to the secondary attitudes instrument, which was a bespoke measure designed to be quick and cost effective to administer. The instrument was not specifically validated prior to its use, although previous research had validated similar instruments using similar items. Nevertheless, the finding about pupil attitudes should be treated with caution.

A final limitation relates to the effect on pupils in receipt of FSM. Although this effect was robust across a range of models, the finding should be treated with some caution due to the relatively small number of FSM pupils.

Future research and publications

There is good evidence, from this and the previous evaluation (Rutt et al., 2014), that Catch Up® Numeracy is a low cost intervention that is attractive to schools. Both this and the previous trial (Rutt et al., 2014) suggest that Catch Up® Numeracy is no more effective than alternative support by TAs. Further research is needed to understand how to maximise the benefits of TA one-to-one support.

The Catch Up® Numeracy intervention seeks to influence TA numeracy support for low attaining pupils in three broad ways: first, Catch Up® provides training for TAs and school coordinators; second, the intervention aims to influence the content and pedagogy of TA sessions by taking a componential approach to enable support tailored to the needs of individual pupils; and, third, the structure of regular, short, logged support sessions is intended to ensure that TA support does take place in a coherent and planned manner. The results of this and the previous trial (Rutt et al., 2014) suggest that the organisational structure and support for TA numeracy support may be more important factors in improving pupil attainment than either the training or the componential structure. However, as already noted, the research itself may have had an impact on sustaining the delivery of TA support by control schools for the duration of the trial because the evaluation team contacted schools on a half-termly basis to ask schools to complete numeracy support session logs. More research is needed to understand how to support schools to sustain regular TA support over time, and to understand how to train TAs in numeracy support.

The process evaluation has highlighted that schools appear to have a strong preference for delivering numeracy support in group rather than one-to-one sessions, and that beyond the trial, many schools are likely to deliver Catch Up® Numeracy support to groups of pupils, perhaps through longer sessions. Given that the Catch Up® Numeracy intervention appeared to be attractive to schools, as well as the findings of the IPE, it is likely that schools would have a preference for group over individual delivery for any intervention. Hence, further research is needed to understand the effects of the composition of delivery and TA support in general. Indeed, the findings of Pellegrini et al.'s (2018) meta-analysis suggest that there would be good justification for trials investigating the effects of group-delivered TA interventions in numeracy.

Finally, the evaluation highlighted a substantial level of pupil mis-selection. Whilst the compliance analysis indicated that this mis-selection was unlikely to have had an effect on the results of this evaluation, this nevertheless raises an issue for future trials involving low attaining pupils, and for interventions aimed at this group more generally. Schools require guidance, support, and training on the identification and selection of pupils.

References

Abayomi, Gelman and Levy (2008) 'Diagnostics for multivariate imputations', *RSS Applied Statistics Series C,* 57 (3), pp. 273–291.

Blatchford, P., Bassett, P., Brown, P., Koutsoubou, M., Martin, C., Russell, A., Webster, R. and Rubie-Davies, C. (2009) 'The impact of support staff in schools. Results from the Deployment and Impact of Support Staff (DISS) project, (Strand 2 Wave 2).

Blatchford, P., Bassett, P., Brown, P. and Webster, R. (2009) 'The effect of support staff on pupil engagement and individual attention', *British Educational Research Journal,* 35 (5), pp. 661–686.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*, 3 (2), pp. 77–101.

Brown, M., Askew, M., Hodgen, J., Rhodes, V., Millett, A., Denvir, H. and Wiliam, D. (2008) 'Individual and cohort progression in learning numeracy ages 5-11: Results from the Leverhulme 5-year longitudinal study', in A. Dowker (ed.), *Mathematical Difficulties: Psychology and Intervention*, Oxford: Elsevier (pp. 85–108).

Browne, W. J. (2012) *MCMC Estimation in MLwiN, v2.26*, Centre for Multilevel Modelling, University of Bristol.

Catch Up® (2017) *Catch Up® Numeracy file v46, 30.11.2017*, Thetford, U.K.: Catch Up®.

Department for Education (2017) 'School workforce in England: November 2016', SRF 25/2017, London: DfE.

Dong, Y. and Peng, C-Y. J. (2013) 'Principled missing data methods for researchers', *SpringerPlus* 2 (222).

Dowker, A. (2009) *What works for children with mathematical difficulties?* Nottingham: DCSF Publications.

Dowker, A. and Sigley, G. (2010) 'Targeted interventions for children with arithmetical difficulties', *Understanding Number Development and Difficulties, BJEP Monograph Series II,* 7, pp. 65–81.

Dowker, A., Bennett, K. and Smith, L. (2012) 'Attitudes to Mathematics in Primary School Children', *Child Development Research*.

Dowker, A., Sarkar, A. and Looi, C. Y. (2016) 'Mathematics Anxiety: What Have We Learned in 60 Years?', *Frontiers in Psychology*, 7 (508).

EFF (2016) *Guidance on cost evaluation*, London: EEF.

EEF (2018) *Evaluation template report 2018*, London: EEF.

Francis, B., Archer, L., Hodgen, J., Pepper, D., Taylor, B. and Travers, M.-C. (2017) 'Exploring the relative lack of impact of research on "ability grouping" in England: a discourse analytic account', *Cambridge Journal of Education*, 47 (1), pp. 1–17.

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

Gelman, A., Hill, J. and Yajima, M. (2012) 'Why We (Usually) Don't Have to Worry About Multiple Comparisons', *Journal of Research on Educational Effectiveness*, 5.2, pp. 189–211.

Gill, J. (2015) *Bayesian Methods: A Social and Behavioral Sciences Approach* (3rd edn), Boca Raton, FL: CRC Press.

Gillham, B. and Hesse, K. (2001) *Basic Number Screening Test: National Numeracy Strategy Edition: Forms A and B, for Ages 7 to 12 Years* (3rd edn), London: Hodder Education.

GL Assessment (2015) 'PT Maths: Technical Information', available at: http://www.gl-assessment.co.uk/sites/gl/files/images/PTM%20Technical%20Information_incl%2011T_FINAL.pdf

Holmes, W. and Dowker, A. (2013) 'Catch Up Numeracy: a targeted intervention for children who are low-attaining in mathematics', *Research in Mathematics Education*, 15 (3), pp. 249–265.

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

Krinzinger, H., Kaufmann, L., Dowker, A., Thomas, G., Graf, M., Nuerk, H.-C. and Willmes, K. (2007) 'Deutschsprachige Version des Fragebogens für Rechenangst (FRA) für 6- bis 9-jährige Kinder', *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie,* 35 (5), pp. 341–351.

Kruschke, J. K. and Liddell, T. M. (2018) 'The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective', *Psychonomic Bulletin and Review*, 25 (1), pp. 178–206.

McBee, M. (2010) 'Modeling Outcomes With Floor or Ceiling Effects: An Introduction to the Tobit Model', *Gifted Child Quarterly*, 54 (4), pp. 314–320.

McDonald, J. H. (2014) *Handbook of biological statistics* (3rd edn), Baltimore, ML: Sparky House Publishing.

MacIntyre, H. and Ireson, J. (2002) 'Within-class ability grouping: placement of pupils in groups and self-concept', *British Educational Research Journal*, 28 (2), pp. 249–263.

Núñez-Peña, M. I., Guilera, G. and Suárez-Pellicioni, M. (2014) 'The Single-Item Math Anxiety Scale: An Alternative Way of Measuring Mathematical Anxiety', *Journal of Psychoeducational Assessment*, 32 (4), pp. 306–317.

Pellegrini, M., Lake, C., Inns, A. and Slavin, R. E. (2018) 'Effective programs in elementary mathematics: A best-evidence synthesis', paper presented at the Annual Meeting of the Society for Research on Educational Effectiveness, Washington, DC.

Rutt, S, Easton, C. and Stacey, O. (2014) 'Catch Up® Numeracy: Evaluation Report and Executive Summary', London: Education Endowment Foundation.

Sharples, J. (2016) 'Six of the best: how our latest reports can help you support teaching assistants to get results', retrieved from: https://educationendowmentfoundation.org.uk/news/six-of-the-best-how-our-latest-reports-can-help-you-support-teaching-assist/

Sharples, J., Webster, R. and Blatchford, P. (2015) 'Making best use of teaching assistants: Guidance Report', London: Education Endowment Foundation.

Slavin, R. E. (2018) 'New Findings on Tutoring: Four Shockers', retrieved from: https://robertslavinsblog.wordpress.com/tag/research-review

Snijders, T. A. B. and Bosker, R. (2011) *Multilevel Analysis: An Introduction To Basic And Advanced Multilevel Modeling*, London: Sage.

Tabachnick, B. G. and Fidell, L. S. (2012) *Using multivariate statistics*, Needham Heights, MA: Allyn and Bacon.

Thomas, G. and Dowker, A. (2000) 'Mathematics anxiety and related factors in young children', *Proceedings of the British Psychological Society Developmental Section Conference*, Bristol, UK.

Tymms, P. (2004) 'Effect sizes in multilevel models', in Schagen, I. and Elliot, K. (eds), *But what does it mean? The use of effect sizes in educational research*, pp. 55–66.

Warhurst, C., Nickson, D., Commander, J. and Gilbert, K. (2013) '"Role stretch": assessing the blurring of teaching and non-teaching in the classroom assistant role in Scotland', *British Educational Research Journa*l, 40 (1), pp. 170–186.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011) 'mice: Multivariate Imputation by Chained Equations in R', *Journal of Statistical Software*, 45 (3), pp. 1–67.

Appendix A: EEF cost rating

Cost ratings are based on the approximate cost per pupil per year of implementing the intervention over three years. More information about the EEF's approach to cost evaluation can be found **here**. Cost ratings are awarded as follows:

Appendix A, Table 16

Cost rating	Description
£ £ £ £ £	*Very low:* less than £80 per pupil per year.
£ £ £ £ £	*Low:* up to about £200 per pupil per year.
£ £ £ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.

Appendix B: Security classification of trial findings

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**Power**	**Attrition**[12]					
5 🔒	Well conducted experimental design with appropriate analysis	MDES < 0.2	0-10%					
4 🔒	Fair and clear quasi-experimental design for comparison (e.g. RDD) with appropriate analysis, or experimental design with minor concerns about validity	MDES < 0.3	11-20%	4		Adjustment for Balance **[-1]**		
3 🔒	Well-matched comparison (using propensity score matching, or similar) or experimental design with moderate concerns about validity	MDES < 0.4	21-30%		→		→	
2 🔒	Weakly matched comparison or experimental design with major flaws	MDES < 0.5	31-40%			Adjustment for threats to internal validity **[-1]**		2
1 🔒	Comparison group with poor or no matching (E.g. volunteer versus others)	MDES < 0.6	41-50%					
0 🔒	No comparator	MDES > 0.6	over 50%					

- **Initial padlock score:** lowest of the three ratings for design, power and attrition. This was a well-randomised experiment, with MDES=0.17 and attrition=17% = [4] padlocks.
- **Reason for adjustment for balance** (if made): [-1] padlocks. A large imbalance was found in the pre-test between both treatment arms, ES=0.21. However, this was unexpectedly large as the imbalance found for KS1 (not used as a pre-test) was ES=0.06. The level of imbalance found in the pre-test is very large enough to compromise the result and would be a reason to reduce 2 padlocks. However, as the balance in other variables is reasonably better and the pre-test was controlled for in a regression model we recommend reducing only [1] padlock. Note, however, that other problems with the outcome test are referred in the threats to validity.
- **Reason for adjustment for threats to validity** (if made): [-1] padlocks. The main threat to the validity of the trial stems from the problems found with the outcome measure. First, large

[12] Attrition should be measured at the pupil level (even for clustered trials) and from the point of randomisation to the point of analysis.

imbalances were found as referred above. Second, floor effects were apparent in the pre- and post-tests (the test was too difficult for some pupils). This is a reason to drop [-1] padlock. In addition, the programme was not delivered to its target group as only around half of the pupils included fulfilled the eligibility criteria to be in the lowest 15-20% of the attainment distribution. Schools generally struggle to identify the lowest attainment pupils so this would be a problem with any targeted intervention (Francis et al 2017). A per-protocol analysis of those correctly identified showed very similar results so no further adjustments for validity are recommended.

- **Final padlock score:** initial score adjusted for balance and internal validity = [2] padlocks

Appendix C: Memorandum of understanding with participating schools

Memorandum of Understanding

Agreement to participate in the Evaluation of Catch Up® Numeracy

Please sign both copies, retaining one and returning the second copy to Ann Fletcher, Catch Up®, Keystone Innovation Centre, Croxton Road, Thetford, IP24 1JD

School Name:

_____

Aims of the Evaluation

The aim of this project is to evaluate the impact of Catch Up® Numeracy, a structured one-to-one intervention which was developed through rigorous academic research and is based on 10 key components of numeracy, on children's numeracy. The results of the research will contribute to our understanding of what works in raising the pupils' attainment and will be widely disseminated to schools in England.

Ultimately, we hope that the evaluation will equip school staff with the knowledge to better support children who struggle with numeracy.

The Project

To test and evaluate, within a rigorous and high quality research framework, the impact on a group of 12 underperforming Year 4 and Year 5 pupils per school, of the Catch Up® Numeracy intervention, when it is delivered by trained teaching assistants for a period of up to 3 terms from September 2016 to July 2017; and to test and evaluate 'in-house approaches' to supporting similar pupils when delivered by teaching assistants for an equivalent time.

The Structure of the Evaluation

Each participating school will be randomly assigned by the external evaluator to the Catch Up® Numeracy group or to the 'active control group', with 75 schools in each group. N.B. There is no 'pure control' group of schools or pupils and all participating pupils will receive support with numeracy difficulties. Pupils will be pre- and post-tested using a standardised test and each participating pupil will undertake an attitudinal assessment. Following completion of the project in 2017, the schools that have not had any staff trained to use Catch Up® Numeracy will be offered Catch Up® Literacy training and support for three staff.

The evaluation is being conducted by a team of external evaluators led by Professor Jeremy Hodgen from the School of Education, University of Nottingham. Schools that are selected and agree to take part are randomly allocated to either the intervention group or the time-equivalent, active control group.

 The schools in the intervention group receive Catch Up® Numeracy training for 2 TAs over a period of 4 x half days, to provide one-to-one support for the pupils for 30 minutes per week for up to 3 terms. A third member of staff, who will undertake the role of Catch Up® Coordinator, will also receive Catch Up® Numeracy training.

 The schools in the active control group receive funding to develop training to enable 2 TAs to provide one-to-one support for the pupils for 30 minutes per week for up to 3 terms.
 All pupils in the evaluation will be tested using a standardised mathematics test and an attitudinal assessment. The external evaluators will undertake a process evaluation, which will include questionnaires and interviews with participating Teaching Assistants and Project Coordinators.

Random allocation is essential to the evaluation as it is the best way of outlining what effect Catch Up® Numeracy has on children's attainment. It is important that schools understand and consent to this process.

Use of Data

Pupils' test responses and any other pupil data will be treated with the strictest confidence. The responses will be collected on paper and accessed by the University of Nottingham evaluation team. Named data will be matched with the National Pupil Database and shared with Catch Up® and the EEF. No individual school or pupil will be identified in any report arising from the research.

Responsibilities

Catch Up® will:
 Deliver the Catch Up® Numeracy integrated training, resource and support package to three members of staff in the intervention group, including providing ongoing support in delivering Catch Up® Numeracy
 Be the first point of contact for any questions about the evaluation
 Provide ongoing support to the school
 Send out regular updates on the progress of the project through a newsletter

The University of Nottingham evaluation team will:
 Conduct the random allocation
 Collect and analyse all the data from the project
 Ensure all staff carrying out assessments are trained and have received CRB clearance
 Provide head teachers with all attainment data after the tests have been completed
 Disseminate research findings

The School will:
 Consent to random allocation and commit to the outcome (whether allocated to the intervention group or the active control group of schools).
 Allow time for each testing phase and liaise with the evaluation team to find appropriate dates and times for testing to take place
 Release two Teaching Assistants and a third member of staff, who will be the school's Project Coordinator, so that they can attend the 4 x half day training sessions.
 Ensure the shared understanding and support of all school staff regarding the project and personnel involved.
 Enable the TAs to support pupils in accordance with the training provided
 Be a point of contact for parents / carers seeking more information on the project.

We commit to the Evaluation of Catch Up® Numeracy as detailed above.

Head teacher [NAME]: _____

Other relevant school staff [NAME]: _____ (Project Coordinator)

Date:_____

Appendix D: Consent and information letters

Information Letter: Catch Up® Numeracy Evaluation

Dear Pupil,

I am writing to let you know about a research project that will be taking place in your school this year. Your teacher will have explained that the project is looking at how to improve the numeracy skills of children. The headteacher of your school has given permission for the school to take part, and the study will involve students in Year 3, 4 and 5.

I would be very grateful for your help in the research, which aims to help teachers, teaching assistants and schools to teach maths in better ways. If you agree to take part, you will be asked to take a couple of maths tests and to fill out a questionnaire. The results of the test will help us with our research but they will not count towards any school assessments, so you do not need to worry about how you do in the test.

The recordings and written copies of interviews will be kept for research purposes. We will share data with our partner researchers at the University of Oxford and at the charity Catch Up®. We will write reports about the research for conferences, teachers and other researchers. We will not use your real name in any research reports and it will not be possible to identify you from our reports.

We would be very grateful if you agree to help us in this project and allow us to collect these data. You should only participate if you want to and choosing not to take part will not disadvantage you in any way.

If you have any questions about the study please ask your teacher or contact Dr. Michael Adkins on 0115 951 4487, or email education_evaluations@nottingham.ac.uk

Yours faithfully,

Jeremy Hodgen

Principal Investigator: Catch Up® Numeracy Evaluation

Professor of Mathematics Education

University of Nottingham

jeremy.hodgen@nottingham.ac.uk

CONSENT FORM FOR PUPILS

Title of Study: Catch Up® ® Numeracy Evaluation

ETHICS REFERENCE: **2016/989/CD**

Thank you for thinking about taking part in this research. The person organising the research will tell you about the project before you agree to take part. If you have any questions about the project, please ask your teacher before you decide whether to join in. You will be given a copy of this Consent Form to keep and look at any time.

Please only complete this form after you have read the Information Letter and listened to an explanation of the research.

- I have read and understood the Information Sheet for the above research project. I have been told why the research is happening and what it involves. I understand and agree to take part.
- I understand that I can choose whether to take part and that I can stop taking part at any time without saying why. I understand that this will not affect what happens to me now or in the future in school. I understand that I can remove any information about me from the study until 31st December 2017.
- I agree to information about me being used in the project in the way explained and as allowed by the law. I understand that any information about me will be kept confidential; I will not be named or identified in any reports from the project.
- I consent to being observed during lessons and taking part in interviews. I understand that these lessons may be video- or audio-recorded and a written record made of what was said. I understand that this information will be stored securely.
- I understand and agree that for the purpose of the study, information provided will be linked with the National Pupil Database (held by the Department for Education), other official records, and shared with Catch Up® ®, the Department for Education, Education Endowment Foundation (EEF), EEF's data contractor FFT Education and in an anonymised form to the UK Data Archive.

If you believe that this study has harmed you in any way or if you wish to make a complaint about the study you can contact the Research Ethics Coordinator of the School of Education, University of Nottingham: **educationresearchethics@nottingham.ac.uk**

If for any reason you wish to opt out of the research, please return this slip to:

_____ _____ _____

Name of Child **Date** **Signature of Parent/Guardian**

LETTER TO PARENTS

Dear Parent/Carer,

Catch Up® Numeracy Evaluation

I am writing to let you know about the Catch Up® Numeracy evaluation, a primary level numeracy intervention being independently evaluated by the University of Nottingham that will be taking place in your child's school this year. The project is aiming to bring children who may be struggling in mathematics to an age appropriate level through one-to-one coaching with a teaching assistant for up to one academic year. The Catch Up® Numeracy intervention and evaluation are funded by the Education Endowment Foundation and the headteacher of your child's school has agreed to take part, and the study will involve pupils in Year 3, 4 and 5.

If your child agrees to take part in the study, your child will be asked to take a couple of maths tests and a questionnaire. The results of the test will help us with our research but they will not count towards any school assessments, so your does not need to worry about how they do in the test.

We would be very grateful if your child agrees to help us in this project and allow us to collect these data. Your child should only participate if he/she wants to; choosing not to take part will not disadvantage your child in any way.

I do hope that we have your support for this important study and that you are happy for your child's data to be used in this way. If you would prefer we did not use your child's data in this research study, please contact the school to opt out. You can do this up to one month after it has been collected.

If you have any questions about the study please ask your teacher or contact Dr. Michael Adkins on 0115 951 4487, or email educational_evaluations@nottingham.ac.uk

Yours faithfully,

Jeremy Hodgen

Principal Investigator: Catch Up® Numeracy Evaluation

Professor of Mathematics Education

University of Nottingham

jeremy.hodgen@nottingham.ac.uk

If for any reason you wish to your child to opt out of the research, please return this slip to:

_____ _____ _____

Name of Child **Date** **Signature of Parent/Guardian**

INFORMATION SHEET FOR TEACHERS

Catch Up® Numeracy Evaluation

We would like to invite you to participate in the evaluation of the Catch Up® Numeracy, an intervention being independently evaluated by the University of Nottingham.

Low attainment in numeracy is a very significant problem in England. Although standards in primary numeracy have risen over the past two decades, overall performance remains below that of the highest achieving jurisdictions and primary performance is not sustained into secondary education (Mullis et al, 2012). However, in contrast to this overall trend, the performance of the very lowest attaining students in England appears to have fallen, which has led to a widening of the attainment gap (Brown et al, 2008).

We have substantial evidence from multiple studies that Catch Up® Numeracy, and the use of the Education Endowment Foundation teaching assistant guidance, does, on average, lead to significant improvements in the attainment of children who struggle with numeracy compared to those in a "business-as-usual" control group. The aim of this evaluation is to see which intervention provides the strongest improvements amongst the participants.

We are focusing on low attaining students in Year 3, 4 and 5. If you agree to take part in the research, we would like 12 of your pupils from the school to sit a test. At the end of the intervention, we would like these pupils to sit another test to understand their improvement, and to fill out an attitudes and anxiety questionnaire.

In addition, we will interview a small sample of teaching assistants and teachers about the students, their perceptions of the students' mathematical understandings and attitudes, their views as to why the students performed poorly on the tests, the curriculum followed, and teaching approaches used. These interviews and lessons observations may be audio-recorded and transcribed. These recordings and transcripts will be shared amongst the research team. We will write reports about the research for conferences, teachers and other researchers. We will not use your real name in these research reports and it will not be possible to identify you from these research reports.

This intervention will take place from September 2016 until July 2017. We would be very grateful if you agree to take part in this project and allow us to collect these data. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. You may withdraw your data from the project up to one month after it has been collected.

If you would like more information or if you have any concerns, please ask one of the researchers or contact: Dr. Michael Adkins on 0115 951 4487, or email: education_evaluations@nottingham.ac.uk

If this study has harmed you in any way or if you wish to make a complaint about the conduct of the evaluation you can contact the Research Ethics Coordinator of the School of Education, University of Nottingham: educationresearchethics@nottingham.ac.uk

Thank you for reading this information sheet and for considering taking part in this research.

Jeremy Hodgen
Principal Investigator: Catch Up® ® Numeracy Evaluation
Professor of Mathematics Education
University of Nottingham
jeremy.hodgen@nottingham.ac.uk

CONSENT FORM FOR TEACHING ASSISTANTS/SCHOOL COORDINATORS

Title of Study: Catch Up® ® Numeracy Evaluation

ETHICS REFERENCE: 2016/989/CD

Thank you for considering taking part in this research. The person organising the research must explain the project to you before you agree to take part. If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time. Please only complete this form after you have read the **Information Sheet** and/or listened to an explanation about the research.

- I have read and understood the Information Sheet for the above research project. The nature and purpose of the research project has been explained to me. I understand and agree to take part.

- I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason. I understand that this will not affect my status now or in the future. Furthermore, I understand that I will be able to withdraw my data from the project up to one month after it has been collected.

- I consent to the processing of my personal information for the purposes explained to me. I understand that such information will be handled in accordance with the terms of the UK Data Protection Act 1998. I understand that confidentiality and anonymity will be maintained and it will not be possible to identify me in any publications.

- I consent to being audio-recorded during interviews and I understand that these data will be stored securely.

- I agree to be contacted in the future by University of Nottingham researchers who would like to invite me to participate in follow up studies to this project.

- I agree that the research team may use my data for future research and understand that any such use of identifiable data would be reviewed and approved by a research ethics committee.

- I understand and agree that for the purpose of the study, information provided will be linked with the National Pupil Database (held by the Department for Education), other official records, and shared with Catch Up®, the Department for Education, Education Endowment Foundation (EEF), EEF's data contractor FFT Education and in an anonymised form to the UK Data Archive.

If you believe that this study has harmed you in any way or if you wish to make a complaint about the study you can contact the Research Ethics Coordinator of the School of Education, University of Nottingham: **educationresearchethics@nottingham.ac.uk**

_____ _____ _____

Name of Teaching Assistant

(or School Coordinator) **Date** **Signature**

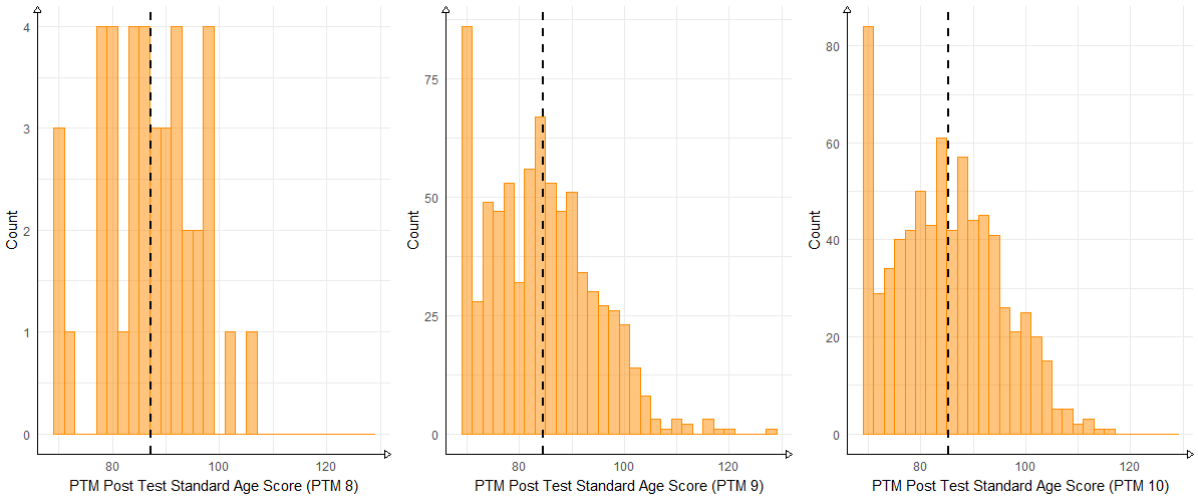
_____ _____ _____

Name of Researcher **Date** **Signature**

Appendix E: Primary outcome modeling

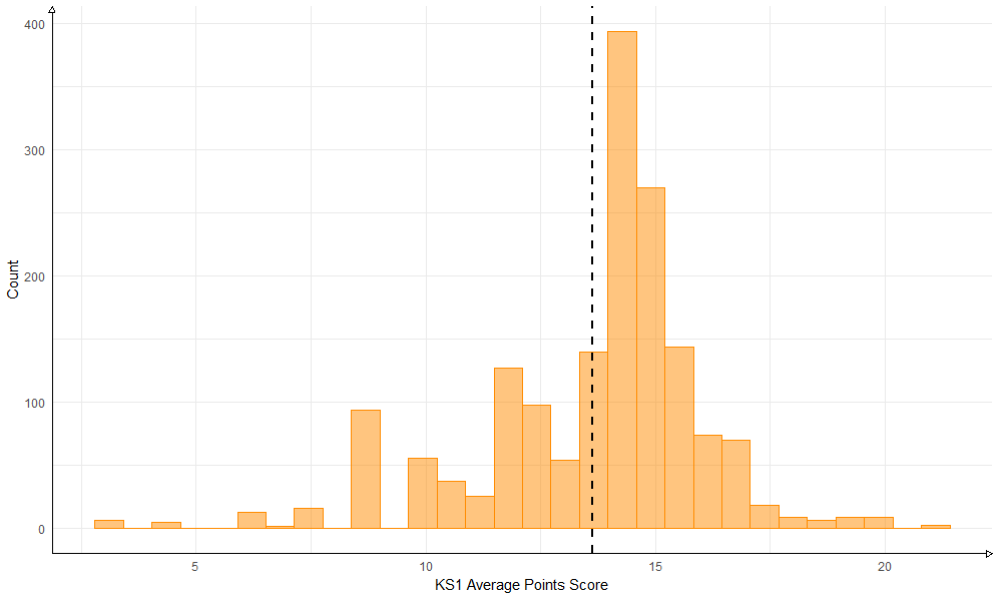
Appendix E: Table 1: Summary of mean differences and effect size of the PTM age adjusted scores by year group

Outcome	Raw means				Effect size	
	Catch Up® group		Active Control group			
	n (missing)	Mean (95% Confidence Interval)	n (missing)	Mean (95% Confidence Interval)	n in model (intervention; control)	Bayesian Hedges g (95% Credible Interval)
Post-test (PTM) Year 4[13]	418 (52)	85.1	418 (59)	84.1	725 (366, 359)	-0.07 (-0.27,0.12)
Post-test (PTM) Year 5	341 (66)	86.1	357 (47)	84.8	698 (341,357)	0.04 (-0.17, 0.26)

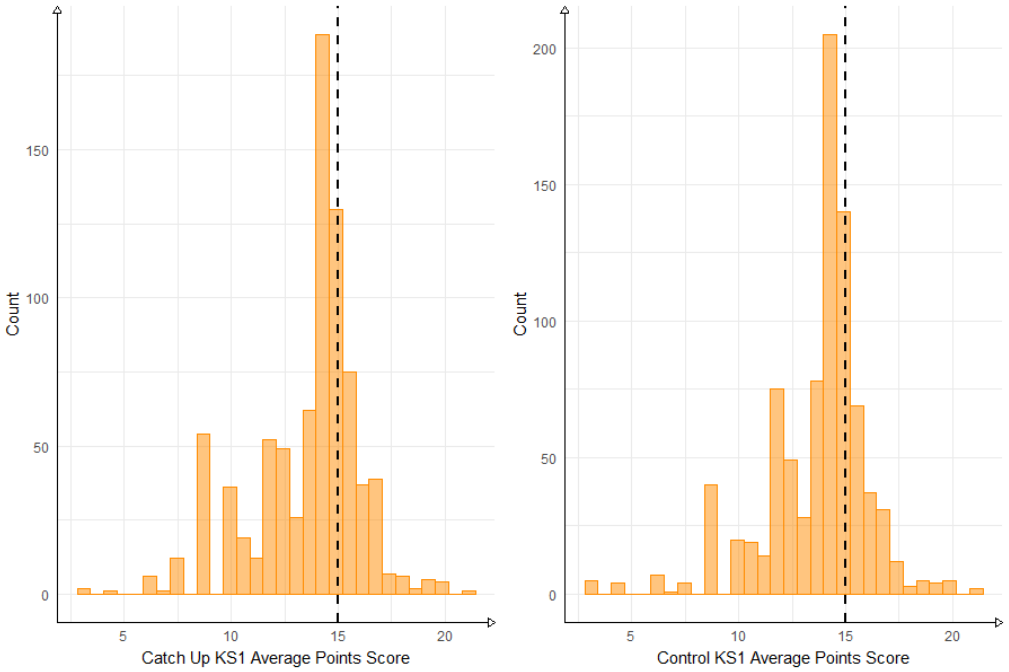


Appendix E: Figure 1: Histograms of the PTM post-test age-adjusted scores by test taken

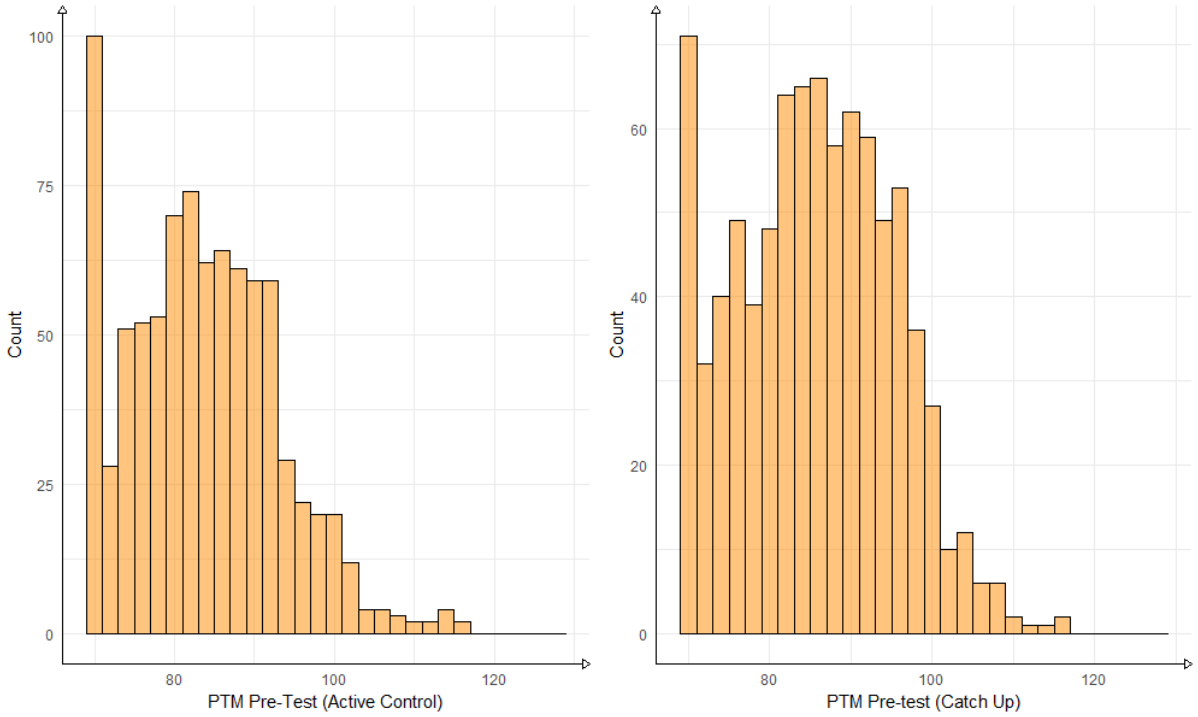
[13] Year 3's were excluded from the analysis due to a very small sample size.



Appendix E: Figure 2: Distribution of pupil-level KS1 average points score



Appendix E: Figure 3: Distribution of pupil-level KS1 average points score by treatment group



Appendix E: Figure 4: Distribution of PTM Pre-Test by treatment group

Appendix E: Table 2: Descriptive quantiles of the prior attainment measures for all pupils

	Quantiles of the Prior Ability Measure Distributions				
Measure	**0%**	**25%**	**50%**	**75%**	**100%**
KS1 Average Points Score	3.0	12.5	14.3	15.0	21.0
PTM Pre-test National Percentile Rank	2	7	16	30	87
PTM Pre-Test Overall Stanine	1	2	3	4	7

Appendix E: Table 3: Bayesian Leave-One-Out Cross-Validation (LOO-CV) model fit statistics for the main model of the primary outcome

Model	LOOIC (SE)	Difference (SE)
Null Model	11116.48 (50.71)	-
Multilevel Null Model	10628.66 (58.68)	-487.82 (40.56)
Varying Intercept with Treatment indicator	10629.40 (58.87)	-0.74 (1.42)
Varying Intercept Model with Treatment indicator and mean centred pre-test variable	9810.08 (54.51)	-818.79 (52.23)

Appendix E: Table 4: Bayesian R^2 statistics based on hypothetical future data with 95% credible intervals for the main model of the primary outcome

Model	Bayes R^2 (SE)	Lower 95% Interval	Upper 95% Interval
Null Model	0.00 (0.00)	0.00	0.00
Multilevel Null Model	0.37 (0.02)	0.33	0.41
Varying Intercept with Treatment indicator	0.37 (0.02)	0.33	0.41
Varying Intercept Model with Treatment indicator and mean centred pre-test variable	0.64 (0.01)	0.62	0.66

Appendix E: Table 5: Multilevel regression coefficients of main model of the primary outcome

Multilevel Unstandardised Regression Coefficients				
Parameter	**Posterior Mean**	**Posterior Standard Deviation**	**Lower 95% Credible Interval**	**Upper 95% Credible Interval**
β_0 Intercept	84.9	1.2	82.4	87.5
β_1 Treatment	-0.3	0.7	-1.7	1.0
β_2 Pre-Test	0.7	0.0	0.7	0.8
σ^2 Randomisation Group	6.0	17.2	0.0	38.7
σ^2 School	6.6	2.3	2.3	11.4
σ^2 TA	10.7	2.3	6.8	15.6
σ^2	38.4	1.6	35.4	41.7

Appendix E: Table 6: Classically estimated ICCs calculated on the basis of complete case analysis for the main model of the primary outcome

	Intra Cluster Correlation		
Test	**TA-Level**	**School-Level**	**Region-Level**
Post-Test Design	0.05	0.10	-
Post-Test Primary Analysis without Imputation	0.10	0.27	0.01
PTM Pre-test	0.15	0.25	0.04
KS1 Average Points Score	0.10	0.23	0.00

Appendix E: Table 7: Classically estimated model fit statistics based on the reduction in the log likelihood for the main model of the primary outcome

Model	Log Likelihood (df)	Difference
Null Model	-5556.391 (2)	-
Multilevel Null Model	-5386.964 (5)	- 169.427
Varying Intercept with Treatment indicator	-5385.077 (6)	- 1.887
Varying Intercept Model with Treatment indicator and mean centred pre-test variable	-4965.057(7)	- 420.02

Appendix E: Table 8: Classically fit primary outcome multilevel regression coefficients for the main model of the primary outcome

Model	Estimate (SE)	Lower 95% Interval	Upper 95% Interval
β_0 Intercept	84.88 (0.63)	82.41	87.50
β_1 Treatment	-0.31 (0.67)	-1.65	0.98
β_2 PTM Pre-test	0.73 (0.02)	0.69	0.77
σ^2 Randomisation Group	0.0 (0.0)	0.00	0.00
σ^2 School	6.6 (2.2)	2.3	10.9
σ^2 TA	10.3 (2.2)	6.0	14.6
σ^2	38.3 (1.6)	35.2	41.4

Appendix E: Table 9: Classically fit primary outcome effect sizes

Model	Effect Size Mean	Lower 95% Interval	Upper 95% Interval
Treatment	-0.04	-0.22	0.13
PTM Pre-test	0.10	0.09	0.10

Appendix E: Table 10: Logistic Regression drop-out model of missingness

Model	Posterior Mean (SE)	Lower 95% Credible Interval	Upper 95% Credible Interval
Intercept	-3.04 (0.45)	-3.92	-2.11
Treatment	0.26 (0.33)	-0.37	0.91
PTM Pre-test	0.00(0.01)	-0.02	0.03
everFSM	0.82 (0.21)	0.41	1.22
Pupil KS1 Score	-0.08 (0.04)	-0.16	0.00
School KS1 Score	0.14 (0.15)	-0.16	0.43
School KS2 Score	-0.02 (0.02)	-0.05	0.01
σ^2 Randomisation Group	0.58 (2.06)	0.00	4.36
σ^2 School	1.96 (0.52)	1.12	3.15
σ^2 Teaching Assistant	0.15 (0.17)	0.00	0.62

Appendix E: Table 11: Primary outcome - Bayesian imputation using a missing data sub-model

Model	Posterior Mean (SE)	Lower 95% Credible Interval	Upper 95% Credible Interval
Intercept	85.0 (1.3)	82.6	87.6
Treatment	-0.1 (0.7)	-1.4	1.2
PTM Pre-test	0.7 (0.0)	0.7	0.8
σ^2 Randomisation Group	5.6(17.4)	0.0	39.0
σ^2 School	6.5 (2.4)	1.9	11.2
σ^2 Teaching Assistant	10.9 (2.4)	6.9	16.1
σ^2	38.4 (1.5)	35.4	41.5

Appendix E: Table 12: Bayesian censored student-t multilevel regression results

Multilevel Regression Coefficients				
Parameter	**Posterior Mean**	**Posterior Standard Deviation**	**Lower 95% Credible Interval**	**Upper 95% Credible Interval**
β_0 Intercept	84.6	1.3	82.0	87.4
β_1 Treatment	-0.3	0.8	-1.8	1.2
β_2 Pre-test	0.8	0.0	0.7	0.8
σ^2 Randomisation Group	6.9	19.8	0.0	43.6
σ^2 School	8.6	2.8	3.2	14.4
σ^2 TA	12.5	2.7	7.9	18.5
σ^2	41.0	2.0	36.8	45.0

Appendix E: Table 13: Multilevel Regression coefficients for the primary model with school-level covariates

Multilevel Regression Coefficients				
Parameter	**Posterior Mean**	**Posterior Standard Deviation**	**Lower 95% Credible Interval**	**Upper 95% Credible Interval**
β_0 Intercept	85.1	1.4	82.4	88.0
β_1 Treatment	-0.4	0.7	-1.8	1.0
β_2 Pre-test	0.7	0.0	0.7	0.8
B_3 School-level KS1	0.1	0.3	-0.6	0.71
B_4 School-level KS2	0.0	0.0	-0.1	0.1
σ^2 Randomisation Group	7.3	21.7	0.0	49.1
σ^2 School	6.4	2.7	0.0	11.7
σ^2 TA	11.4	2.7	7.0	17.7
σ^2	38.7	1.6	35.7	42.1

Appendix E: Table 14: Multilevel Regression coefficients for the gain score model

Multilevel Regression Coefficients				
Parameter	**Posterior Mean**	**Posterior Standard Deviation**	**Lower 95% Credible Interval**	**Upper 95% Credible Interval**
β_0 Intercept	0.23	1.51	-2.72	3.39
β_1 Treatment	-0.86	0.70	-2.24	0.50
σ^2 Randomisation Group	9.49	22.02	0.02	62.17
σ^2 School	4.46	2.56	0.10	9.71
σ^2 TA	15.07	2.88	9.83	21.02
σ^2	42.23	1.74	39.03	45.86

Appendix E: Table 15: Multilevel Regression coefficients for the FSM subgroup analysis

Multilevel Regression Coefficients				
Parameter	**Posterior Mean**	**Posterior Standard Deviation**	**Lower 95% Credible Interval**	**Upper 95% Credible Interval**
β_0 Intercept	83.83	1.04	81.83	85.91
β_1 Treatment	-0.88	0.81	-2.46	0.69
B_2 Pre-Test	0.69	0.04	0.62	0.76
σ^2 Randomisation Group	3.47	10.23	0.00	26.93
σ^2 School	5.30	3.23	0.13	12.10
σ^2 TA	8.21	3.69	1.27	15.70
σ^2	39.91	3.10	34.20	46.53

Appendix E: Table 16: Effect size estimates for FSM subgroup analysis

Parameter	Effect Size Mean (SE)	Lower 95% Credible Interval	Upper 95% Credible Interval
Treatment	-0.12 (0.11)	-0.33	0.09
PTM Pre-test	0.09 (0.01)	0.08	0.10

Appendix E: Table 17: Comparison of different models of the treatment effect for raw post-test PTM scores for the primary outcome on FSM pupils (using the sample of pupils for which PTM pre-test and KS1 scores were available, N=1395)

Model	Posterior Mean Treatment Effect	Lower 95% Credible Interval	Upper 95% Credible Interval
1: Main model: PTM as pre-test covariate	-0.14	-0.33	0.05
2: KS1 as only pre-test covariate	-0.10	-0.29	0.10
3: PTM and KS1 as pre-test covariates	-0.11	-0.31	0.09

Appendix F: Secondary outcome modelling

Appendix F: Table 1: Multiple outcome model, simultaneously fitting the primary and secondary outcome.

Model	Progress Test In Mathematics	Lower and Upper 95% Credible Interval	Pupil Attitudes Survey to Maths	Lower and Upper 95% Credible Interval
β_0 Intercept	85.3 (0.95)	83.2, 87.2	7.48 (0.19)	7.1, 7.9
β_1 Treatment	-0.4 (0.7)	-1.8, 1.0	0.2 (0.1)	-0.04, 0.44
B_2 Pre-test	0.3 (0.0)	0.3, 0.4	0.8 (0.0)	0.7, 0.8
σ^2 Randomisation Group	3.2 (6.1)	0.00: 20.5	0.1 (0.3)	0.0: 0.8
Slope Correlation	-0.1 (0.6) (-1, 0.9)			
σ^2 School	6.2 (2.9)	0.6, 12.3	0.2 (0.1)	0.0, 0.3
Slope Correlation	-0.2 (0.4) (-0.8, 0.6)			
σ^2 Teaching Assistant	12.0 (3.0)	6.7, 18.7	0.2 (0.1)	0.0, 0.4
Slope Correlation	-0.1 (0.3) (-0.7: 0.4)			
σ^2	36.9 (1.9)	33.4, 40.7	1.9 (0.1)	1.7, 2.1
Correlation	0.1 (0.0) (0.0: 0.2)			

Appendix F: Table 2: Secondary Outcome Multilevel regression coefficients

Multilevel Regression Coefficients				
Model	**Posterior Mean**	**Posterior Standard Deviation**	**Lower 95% Credible Interval**	**Upper 95% Credible Interval**
β_0 Intercept	7.51	0.22	7.15	7.89
β_1 Treatment	0.17	0.12	-0.07	0.41
β_2 Pre-test	0.34	0.03	0.28	0.39
σ^2 Randomisation Group	0.16	0.86	0.00	0.99
σ^2 School	0.19	0.08	0.04	0.36
σ^2 TA	0.15	0.08	0.00	0.33
σ^2	1.95	0.09	1.77	2.14

Appendix F: Table 3: Secondary outcome - Bayesian Imputation using a missing data sub-model

Model	Posterior Mean (SE)	Lower 95% Credible Interval	Upper 95% Credible Interval
β_0 Intercept	7.51 (0.22)	7.19	7.89
β_1 Treatment	0.18 (0.13)	-0.07	0.42
β_2 Pre-test	0.34 (0.03)	0.29	0.39
σ^2 Randomisation Group	0.17 (1.35)	0.00	0.83
σ^2 School	0.19 (0.08)	0.03	0.35
σ^2 Teaching Assistant	0.15 (0.08)	0.01	0.33
σ^2	1.95(0.09)	1.77	2.14

Appendix G: Compliance and dosage modelling

Appendix G: Table 1: Multilevel regression coefficients for primary model with the inclusion of interaction effects of the termly roll-off covariate (monotonic effects)

Multilevel Regression Coefficients				
Parameter	**Posterior Mean**	**Posterior Standard Deviation**	**Lower 95% Credible Interval**	**Upper 95% Credible Interval**
β_0 Intercept	86.2	1.6	83.2	89.3
β_1 Treatment	1.2	1.2	-1.2	3.6
β_2 Pre-test	0.7	0.0	0.67	0.8
B_3 Term support withdrawn	-1.4	0.9	-3.3	0.4
B_4 Term support withdrawn * Treatment	-2.2	1.3	-4.6	0.3
Simplex Parameters:				
Term support withdrawn (Spring)	0.6	0.3	0.1	1.0
Term support withdrawn (Summer)	0.4	0.3	0.0	0.9
Term support withdrawn (Spring) * Treatment	0.6	0.3	0.1	1.0
Term support withdrawn (Summer) * Treatment	0.4	0.3	0.0	0.9
σ^2 Randomisation Group	7.7	25.0	0.0	0.5
σ^2 School	5.9	2.6	0.6	11.3
σ^2 TA	11.3	2.7	6.8	17.4
σ^2	37.6	1.7	34.5	41.1

Appendix G: Table 2: Effect size for primary model with the inclusion of the termly roll-off covariate

Parameter	Posterior Effect Size Mean	Posterior Effect Size Standard Deviation	Lower 95% Credible Interval	Upper 95% Credible Interval	Proportion of posterior >0	Proportion of posterior <0
Treatment (Autumn)	0.16	0.16	-0.15	0.47	0.83	0.17
Treatment (Spring)	-0.02	0.11	-0.23	0.20	0.41	0.59
Treatment (Summer)	-0.05	0.18	-0.41	0.30	0.39	0.61

Appendix G: Table 3: Multilevel regression coefficients for primary model with the inclusion of the one-to-one session covariate

Multilevel Regression Coefficients				
Parameter	**Posterior Mean**	**Posterior Standard Deviation**	**Lower 95% Credible Interval**	**Upper 95% Credible Interval**
β_0 Intercept	85.6	1.4	82.7	88.3
β_1 Treatment	-1.3	1.0	-3.3	0.7
β_2 Pre-test	0.7	0.0	0.7	0.8
β_3 Sessions delivered on a one-to-one basis	-1.3	1.3	-3.8	1.3
σ^2 Randomisation Group	5.4	16.7	0.00	37.3
σ^2 School	8.3	3.1	2.1	14.8
σ^2 TA	10.0	2.8	5.3	16.5
σ^2	39.0	1.9	35.6	42.8

Appendix G: Table 4: Effect size for primary model with the inclusion of the one-to-one session covariate.

Parameter	Posterior Effect Size Mean	Posterior Effect Size Standard Deviation	Lower 95% Credible Interval	Upper 95% Credible Interval	Proportion of posterior >0	Proportion of posterior <0
Treatment	-0.16	0.13	-0.24	0.05	0.11	0.89
Pre-Test	0.09	0.01	0.07	0.10	1	0
One-to-one sessions	-0.16	0.16	-0.47	0.16	0.16	0.84

Appendix G: Table 5: Multilevel regression coefficients for primary model with the inclusion of the one-to-one session covariate

Multilevel Regression Coefficients				
Parameter	**Posterior Mean**	**Posterior Standard Deviation**	**Lower 95% Credible Interval**	**Upper 95% Credible Interval**
Intercept	85.28	1.40	82.46	88.08
Treatment	-0.82	1.03	-2.84	1.18
Pre-Test	0.72	0.02	0.68	0.77
One to one teaching Vs. Group (Baseline category: All the time)				
One to one teaching: Most of the time	0.22	1.40	-2.56	2.95
One to one teaching: Half and half	0.39	1.77	-3.15	3.80
One to one teaching: Occasionally	-0.26	1.54	-3.30	2.68
One to one teaching: Almost Never	-1.33	1.30	-3.83	1.26
σ^2 Randomisation Group	5.54	13.58	0.00	40.72
σ^2 School	6.94	2.77	1.55	12.70
σ^2 Teaching Assistant	11.46	2.65	6.92	17.32
σ^2	38.82	1.76	35.81	42.58

Appendix G: Table 6: Effect size for primary model with the inclusion of the one-to-one session covariate.

Parameter	Posterior Effect Size Mean	Posterior Effect Size Standard Deviation	Lower 95% Credible Interval	Upper 95% Credible Interval	Proportion of posterior >0	Proportion of posterior <0
Treatment	-0.11	0.13	-0.36	0.15	0.22	0.78
Pre-Test	0.09	0.01	0.07	0.10	1	0
One to one teaching: Most of the time	0.02	0.18	-0.33	0.37	0.57	0.43
One to one teaching: Half and half	0.05	0.22	-0.40	0.49	0.40	0.60
One to one teaching: Occasionally	-0.03	0.20	-0.42	0.34	0.43	0.57
One to one teaching: Almost Never	-0.17	0.17	-0.49	0.15	0.86	0.14

Appendix G: Table 7: Multilevel regression coefficients for primary model with the inclusion of the pre-test and treatment interaction

Multilevel Regression Coefficients				
Parameter	**Posterior Mean**	**Posterior Standard Deviation**	**Lower 95% Credible Interval**	**Upper 95% Credible Interval**
β_0 Intercept	84.9	1.3	82.4	87.6
β_1 Treatment	-0.3	0.7	-1.6	1.0
β_2 Pre-test	0.7	0.0	0.7	0.8
β_3 Pre-test * Treatment	0.0	0.0	-0.1	0.1
σ^2 Randomisation Group	6.2	17.0	0.0	43.4
σ^2 School	6.4	2.3	2.1	11.3
σ^2 TA	10.9	2.3	6.9	16.0
σ^2	38.5	1.6	35.4	41.7

Appendix G: Table 8: Multilevel regression coefficients for primary model per protocol

Multilevel Regression Coefficients				
Parameter	**Posterior Mean**	**Posterior Standard Deviation**	**Lower 95% Credible Interval**	**Upper 95% Credible Interval**
β_0 Intercept	84.9	1.3	82.2	87.7
β_1 Treatment	-0.6	0.7	-2.0	0.9
β_2 Pre-test	0.8	0.0	0.7	0.8
σ^2 Randomisation Group	6.4	17.8	0.0	41.8
σ^2 School	4.1	2.8	0.1	10.0
σ^2 TA	13.8	3.3	8.1	20.8
σ^2	35.2	2.0	31.6	39.3

Appendix H: Compliance criteria

Compliance for schools in the intervention group was agreed with the developer and defined as follows:

- Pupil selection: Selected pupils eligible for the intervention and were judged to be struggling with numeracy when selected by their schools. The usual Catch Up® Numeracy selection process includes social, emotional, physical and other characteristics in addition to attainment below chronological age (Catch Up®, 2017), but, for the purposes of the compliance analysis for this trial, the eligible pupils will be considered to be amongst the lowest attaining 15-20% of children for their age [Full compliance required; evidence pre-test age adjusted scores.]
- Adequate staffing: School identified two TAs to deliver the intervention and one school coordinator (and any staff leaving were replaced). [Full compliance required; evidence: initial data provided by school and additional ongoing records collected by Catch Up.]
- Attendance at training: The two TAs and the school coordinator attended all relevant initial training sessions. [Full compliance required; evidence: records collected by Catch Up.]
- Initial assessment: All pupils were assessed according to Catch Up guidance at the start of the intervention. [Full compliance required; evidence: records collected by Catch Up.]
- Termly review: Termly review took place for all pupils in the intervention to assess whether pupils should be “rolled off” the intervention. [Full compliance required; evidence: records collected by Catch Up.]
- Support sessions in school (dosage): Each pupil should receive 2 x 15 minute of one-to-one support sessions for each week that they receive the intervention.

Appendix I: Survey Results

Teaching Assistant Questions	Active Control (N=109)		Catch Up (N=109)	
Where are most of the sessions held?				
In a separate room	49	45%	62	57%
In the classroom	19	17%	0	0%
Another space (e.g. corridor outside the classroom)	33	30%	37	34%
It varies from week to week	8	7%	10	9%
About how often do you talk to the class teacher about Catch Up® (e.g. your pupils' progress/resources/planning)				
Each week	85	78%	49	45%
Once a half term	18	17%	39	36%
Once a term or less	4	4%	15	14%
Never	2	2%	6	6%
About how often do you talk to the co-ordinator about Catch Up® (e.g. your pupils' progress/resources/planning)				
Each week	34	31%	38	35%
Once a half term	46	42%	48	44%
Once a term or less	19	17%	19	17%
Never	8	7%	4	4%
About how often do you talk to other TAs about Catch Up® (e.g. your pupils' progress/resources/planning)				
Each week	59	54%	65	60%
Once a half term	17	16%	22	20%
Once a term or less	16	15%	6	6%
Never	17	16%	15	14%
How long do most Catch Up® sessions last?				
15 minutes	11	10%	59	54%
Over 15 minutes	69	63%	30	28%
Under 15 minutes	9	8%	0	0%
Varies from week to week	19	17%	19	17%

Teaching Assistant Questions	Active Control (N=109)		Catch Up (N=109)	
[For Catch Up schools] I have undertaken the following training (please tick all the appropriate boxes)				
Attended the pre-project Catch Up¨ training sessions			93	85%
Attended one or more Catch Up¨ review sessions			73	67%
Other			4	4%
[For active control schools] I have undertaken the following training (please tick all the appropriate boxes)				
Read a training guide	27	25%		
A few hours face-to-face training sessions	20	18%		
Several face-to-face training sessions	25	23%		
An online course	1	1%		
No training	36	33%		
Other	18	17%		
Missing	1	1%		
For each pupil, about how many sessions are timetabled, per pupil, per week?				
One	29	27%	3	3%
Two	35	32%	85	78%
Over two	27	25%	8	7%
Varies from week to week	18	17%	13	12%
If you have a Maths GCSE/O level, please give the grade.				
*Grade A or A**	10	9%	9	8%
GradeB	25	23%	21	19%
Grade C	29	27%	24	22%
Other	5	5%	7	6%
None	40	37%	48	44%

Teaching Assistant Likert Scale Items	Control Mean (SD)	Catch Up Mean (SD)	*t*	df	*p*	N	*d*
An initial formal assessment informs the content of the sessions / The initial Catch Up assessment for learning informs the content of the sessions[14]	3.5 (1.4)	4.6 (0.7)	-7.3	159.4	0.000	217	1.00
I have sufficient time to prepare for the numeracy sessions / I have sufficient time to prepare for Catch Up	3.5 (1.2)	3.2 (1.2)	2.2	214.9	0.028	217	-0.30
I keep up to date records of pupil progress in the numeracy sessions / The Catch Up documentation is kept up to date e.g. Sessions logs learner profiles	4.3 (1.0)	4.7 (0.5)	-4.3	155.5	0.000	218	0.58
I need to prepare for the numeracy sessions / I need to prepare for the Catch Up numeracy sessions	3.9 (1.1)	4.1 (1.0)	-1.3	211.5	0.193	215	0.18
I use resources that are specifically designed for the numeracy sessions / When delivering Catch Up I use the downloadable Catch Up pupil resources	4.1 (1.1)	3.6 (1.2)	3.5	210.7	0.001	216	-0.48
In the sessions I use the same pupil resources as used in recent maths lessons / In the sessions I use the same pupil resources as used in recent maths lessons	3.2 (1.3)	2.3 (1.2)	5.1	214.4	0.000	218	-0.69
Teacher's assessment of pupils progress in recent lessons informs the session content / Teacher's assessment of pupils progress in recent lessons informs the session content	3.4 (1.5)	2.6 (1.3)	4.6	213.7	0.000	217	-0.62
The numeracy session I deliver follows the same structure from week to week / I strictly follow the session as outlined by Catch Up	4.0 (1.0)	4.4 (0.7)	-3.0	190.1	0.003	218	0.41
The resources I use are engaging for pupils / The downloadable resources are engaging for pupils	4.3 (0.7)	3.5 (1.0)	6.8	191.9	0.000	217	-0.92
When delivering the numeracy sessions they are one to one / When delivering Catch Up the sessions are one to one	3.0 (1.7)	5.0 (0.1)	-12.3	109.4	0.000	218	1.67
As a result of delivering the numeracy sessions I enjoy teaching maths more / As a result of delivering Catch Up I enjoy teaching maths more	3.7 (0.9)	3.4 (1.1)	2.3	203.2	0.024	216	-0.31

[14] The wording differed slightly for the two groups. The active control version is given first.

Teaching Assistant Likert Scale Items	Control Mean (SD)	Catch Up Mean (SD)	*t*	df	*p*	N	*d*
As a result of delivering the numeracy sessions I find maths easier / As a result of delivering Catch Up I find maths easier	3.5 (1.1)	3.1 (1.1)	2.4	214.0	0.016	216	-0.33
As a result of delivering the numeracy sessions I like maths more / As a result of delivering Catch Up I like maths more	3.5 (1.0)	3.2 (1.1)	2.1	210.3	0.035	215	-0.29
As a result of delivering the numeracy sessions my own teaching of maths has improved / As a result of delivering Catch Up my own teaching of maths has improved	4.0 (0.9)	3.7 (1.0)	2.6	209.2	0.009	216	-0.36
Highly structured numeracy sessions are easier to deliver / The structure of Catch Up makes it easy to deliver the sessions	4.1 (0.8)	4.1 (0.8)	-0.3	213.2	0.773	216	0.04
I have the maths skills to deliver the numeracy sessions properly / I have the maths skills to deliver Catch Up properly	4.4 (0.6)	4.5 (0.6)	-0.8	213.8	0.410	216	0.11
I would like to continue with the numeracy sessions next year / I would like to continue with Catch Up next year	3.8 (1.1)	3.7 (1.2)	0.6	209.9	0.551	215	-0.08
My time out to deliver the sessions adversely impacts class management / My time out to deliver Catch Up adversely impacts class management"	2.6 (1.0)	3.2 (1.1)	-3.9	208.9	0.000	216	0.53
Pupils' attitude to maths improves as a result of numeracy sessions / Pupils' attitude to maths improves as a result of Catch Up sessions	4.2 (0.6)	4.1 (0.9)	1.5	191.1	0.147	216	-0.20
Pupils' maths significantly improves as a result of the numeracy sessions / Pupils' maths significantly improves as a result of Catch Up	4.0 (0.7)	3.8 (0.8)	1.7	209.6	0.087	216	-0.23
Pupils recognise that the sessions help them with their maths lessons / Pupils recognise that Catch Up helps them with their maths lessons	4.2 (0.7)	4.1 (0.9)	1.5	200.8	0.124	217	-0.21

School Coordinator Questions	Active Control (N=64)		Catch Up (N=61)	
The pupils have completed an initial numeracy assessment.				
All pupils have completed it	55	86%	56	92%
Some pupils have completed it	6	9%	5	8%
Few or no pupils have completed it	3	5%	0	0%
The decision to end numeracy sessions for a pupil is based on (please tick all appropriate options):				
Professional judgement	22	34%	20	33%
Results of standardised test	10	16%	10	16%
Staff resourcing	6	9%	2	3%
Other	10	16%	1	2%
A combination of the above	26	41%	36	59%
About how often do you talk to the TA about the numeracy sessions?				
Talk each week	25	39%	29	48%
Talk once a half term	33	52%	32	52%
Once a term or less	6	9%	0	0%
Have you observed a session (formally or informally)?				
Yes	40	63%	50	82%
No	24	38%	11	18%
The numeracy sessions are scheduled into the pupils' timetable.				
Yes	61	95%	56	92%
No	3	5%	5	8%
[For Catch Up schools] Regarding TAs' work, I have undertaken the following training/learning (please tick all the appropriate boxes)				
Read the EEF report 'Making best use of teaching assistants'			53	87%
Attended the Catch Up¨ pre-project training sessions			51	84%
Attended one or more Catch Up¨ review sessions			40	66%
Attended an EEF 'Making best use of teaching assistants' guidance event			12	20%
Used the EEF online course 'Making best use of teaching assistants'			9	15%
Other			3	5%
No training			1	2%

School Coordinator Questions	Active Control (N=64)		Catch Up (N=61)	
[For active control schools] Regarding TAs' work, I have undertaken the following training/learning (please tick all the appropriate boxes)				
Read the EEF report 'Making best use of teaching assistants'	51	80%		
Other	19	30%		
Attended an EEF 'Making best use of teaching assistants' guidance event	17	27%		
Participated in the EEF online course 'Making best use of teaching assistants'	5	8%		
No training	4	6%		
Missing	1	2%		

School Coordinator Likert Scale Items	Control Mean (SD)	Catch Up Mean (SD)	*t*	df	*p*	N	*d*
Quality of the sessions is independent of TAs' experience and skills	3.0 (1.2)	3.1 (1.2)	-0.2	122.8	0.818	125	0.04
Records of pupil progress in the numeracy sessions are kept up to date	4.1 (0.8)	4.5 (0.6)	-3.4	121.6	0.001	125	0.61
Highly structured numeracy sessions make it easier for TAs' to deliver them	4.1 (0.7)	4.2 (0.7)	-0.5	122.4	0.642	125	0.08
TAs' time out to deliver the sessions adversely impacts class management	2.8 (1.1)	3.1 (1.1)	-1.5	121.0	0.126	124	0.28
As a result of delivering the numeracy sessions TAs' teaching of maths has improved	3.6 (0.8)	3.8 (0.8)	-1.2	120.9	0.216	125	0.22
Pupils' maths significantly improves as a result of the numeracy sessions	3.9 (0.6)	3.8 (0.8)	0.3	112.6	0.767	125	-0.05
Pupils' attitude to maths improves as a result of the numeracy sessions	3.9 (0.8)	4.2 (0.6)	-1.9	116.0	0.063	124	0.34
The class teacher recognises that the numeracy sessions help pupils with their maths lessons	4.2 (0.5)	3.8 (0.8)	3.0	99.7	0.003	125	-0.54
Pupils recognise that the sessions help them with their maths lessons	4.1 (0.6)	4.1 (0.6)	0.4	122.9	0.682	125	-0.07
I would like to continue with the numeracy sessions next year	3.7 (1.0)	4.1 (0.9)	-2.4	121.4	0.019	125	0.42

School Leadership Team Questions	Active Control (N=49)		Catch Up (N=50)	
About how often do you talk to the TA about the sessions (e.g. pupils' progress/resources)?				
Each week	14	29%	13	26%
Once a half term	23	47%	28	56%
Once a term or less	11	22%	8	16%
Never	1	2%	1	2%
The decision to end the sessions for a pupil is based on (please tick all appropriate options):				
A combination of the above	28	57%	28	56%
Professional judgement	15	31%	13	26%
A standardised test	3	6%	4	8%
Staff resourcing	2	4%	1	2%
A combination of the above	28	57%	28	56%
Other	1	2%	4	8%
Have you observed a session (formally or informally)?				
Yes	34	69%	37	74%
No	15	31%	13	26%
[For Catch Up schools] Regarding TAs' work, I have undertaken the following training/learning (please tick all the appropriate boxes)				
Read the EEF report 'Making best use of teaching assistants'			30	60%
Attended the Catch Up¨ pre-project training sessions			28	56%
Attended an EEF 'Making best use of teaching assistants' guidance event			15	30%
No training			7	14%
Used the EEF online course 'Making best use of teaching assistants'			4	8%
Attended one or more Catch Up¨ review sessions			0	0%
Other			0	0%
[For active control schools] Regarding TAs' work, I have undertaken the following training/learning (please tick all the appropriate boxes)				
Read the EEF report 'Making best use of teaching assistants'	39	80%		
Attended an EEF 'Making best use of teaching assistants' guidance event	16	33%		
Other	14	29%		
Participated in the EEF online course 'Making best use of teaching assistants'	3	6%		
No training	3	6%		

Senior Leadership Team Likert Scale Items	Control Mean (SD)	Intervention Mean (SD)	*t*	df	*p*	N	*d*
Quality of the sessions is independent of TAs' experience and skills	3.0 (1.1)	2.8 (1.4)	0.7	91.3	0.476	98	-0.14
TAs' time out to deliver the sessions adversely impacts class management	2.3 (0.9)	2.6 (0.9)	-1.7	95.9	0.092	98	0.34
As a result of delivering the numeracy sessions TAs' teaching of maths has improved	3.9 (0.8)	3.8 (0.6)	0.4	91.5	0.674	97	-0.09
Pupils' maths significantly improves as a result of the numeracy sessions	3.9 (0.7)	3.8 (0.6)	0.4	94.8	0.696	98	-0.08
Pupils' attitude to maths improves as a result of the numeracy sessions	4.2 (0.6)	4.2 (0.5)	0.1	94.7	0.945	98	-0.01
Highly structured numeracy sessions make it easier for TAs to deliver them	4.3 (0.6)	4.0 (0.5)	2.5	93.2	0.016	97	-0.50
Pupils recognise that the sessions help them with their maths lessons	4.2 (0.6)	4.0 (0.6)	1.9	95.1	0.061	98	-0.38
The class teacher recognises that the numeracy sessions help pupils with their maths lessons	4.2 (0.7)	3.9 (0.6)	2.6	90.4	0.010	98	-0.53
I would like to continue with the numeracy sessions next year	3.9 (1.0)	4.0 (1.0)	-0.3	95.6	0.747	98	0.07
Quality of the sessions is independent of TAs' experience and skills	3.0 (1.1)	2.8 (1.4)	0.7	91.3	0.476	98	-0.14
TAs' time out to deliver the sessions adversely impacts class management	2.3 (0.9)	2.6 (0.9)	-1.7	95.9	0.092	98	0.34

.

This document is available for download at **www.educationendowmentfoundation.org.uk**

Education
Endowment
Foundation

The Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP
www.educationendowmentfoundation.org.uk