

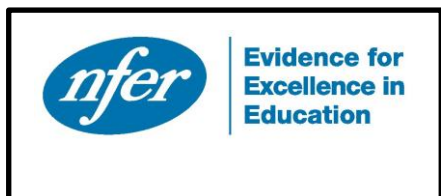


Chatterbooks

Evaluation Report and Executive Summary

May 2014

Independent evaluators:



Dr Ben Styles

Rebecca Clarkson

Katherine Fowler

The Education Endowment Foundation (EEF)

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.

We aim 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;
- Encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

Founded by the education charity the Sutton Trust, as lead charity in partnership with Impetus Trust, the EEF is funded by an initial £125m grant from the Department for Education. With investment and fundraising income, the EEF intends to award as much as £200m by 2026.



Impetus
Private Equity Foundation

Department for
Education

Literacy Catch-up Round

This study was funded by the Education Endowment Foundation as one of 23 projects funded through a themed funding round on literacy catch-up at the primary-secondary transition. Projects funded within this round aimed to identify effective ways to support pupils who do not achieve Level 4 in English by the end of Key Stage 2. It was one of four programmes with a particular focus on comprehension.

For more information please contact:

James Richardson
Senior Analyst
Education Endowment Foundation
9th Floor, Millbank Tower
21-24 Millbank
SW1P 4QP



Education
Endowment
Foundation

p: 020 7802 1924
e: james.richardson@eefoundation.org.uk
w: www.educationendowmentfoundation.org.uk

About the evaluator

The programme was independently evaluated by a team from the National Foundation for Educational Research. The lead evaluator was Dr Ben Styles. For the impact evaluation he was assisted by Sally Bradshaw and Alix Godfrey and for the process evaluation by Rebecca Clarkson and Katherine Fowler.

Contact details:

Dr Ben Styles

The Mere
Upton Park
Slough
Berkshire
SL1 2DQ

p: 01753 637386

e: b.styles@nfer.ac.uk

Contents

Executive summary	2
Introduction.....	5
Methodology	8
Impact evaluation	14
Process evaluation	20
Conclusion	26
References	28
Acknowledgements.....	28
Appendices	28

Executive Summary

The project

Chatterbooks is an extracurricular reading initiative that aims to increase a child's motivation to read by providing schools with tools and resources to encourage reading for pleasure. The intervention developed for this trial consisted of nine weekly one hour sessions where the pupils read and discussed an age-appropriate book. The programme was delivered by trained graduates to Year 7 pupils who had not reached a secure Level 4 in English at the end of Key Stage 2.

Chatterbooks was developed by the Reading Agency as an extracurricular activity, organised in local libraries on Saturday mornings by volunteers. The sessions are intended for six to twelve-year-olds and attendance is voluntary. For the purposes of this evaluation, the Chatterbooks programme was altered significantly to test its impact in a more structured and formalised school environment. A variation called Chatterbooks Plus was delivered alongside the Chatterbooks programme. In this intervention, fifteen minutes of a sixty minute session were replaced with dialogic reading where children read aloud and were offered explicit prompts.



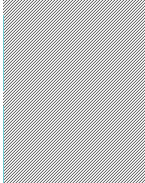
The target population for this evaluation was pupils in secondary schools in an area of the Midlands accessible from Coventry University. The programme was delivered by trained graduates at Coventry University who had received training from the Reading Agency and Professor Clare Wood (for the dialogic reading component). Intervention delivery took place from April to June 2013.

Chatterbooks aims to encourage reading for pleasure. It is assumed that this translates into an improvement in reading ability. This evaluation set out to measure what impact the scheme had on reading ability as measured by the GL Assessment New Group Reading Test.

The evaluation was funded by the Education Endowment Foundation as one of 23 projects focused on literacy catch-up at the primary-secondary transition. It was one of four programmes funded with a particular focus on reading comprehension.

What impact did it have?

There was no evidence of impact of either Chatterbooks or Chatterbooks Plus on pupils' reading ability either immediately after the interventions or at a three-month follow-up. The headline findings suggest that Chatterbooks and Chatterbooks Plus had a slightly negative impact on reading ability with effect sizes of -0.14 and -0.01 respectively. However, they are not statistically significant, suggesting that the difference in outcomes between the control and intervention group occurred by chance. Consequently, the effect sizes are indistinguishable from zero. The results for children on free school meals in both interventions should be interpreted in the same way.

Group	Number of pupils	Effect size (95% confidence interval*)	Estimated months' progress	Is this finding statistically significant?*	Evidence strength**
Chatterbooks	159	-0.14 (-0.31, 0.03)	-2	No	
Chatterbooks Plus	150	-0.01 (-0.18, 0.16)	0	No	
Chatterbooks (FSM pupils)	37	-0.30 (-0.63, 0.03)	-4	No	
Chatterbooks Plus (FSM pupils)	40	-0.09 (-0.39, 0.21)	-1	No	
*Effect sizes with confidence intervals that pass through 0 are not 'statistically significant', suggesting that the difference occurred by chance.					
** Evidence ratings are not provided for sub-group analyses, which will always be less secure than overall findings. For more information about evidence ratings, see Appendix C.					

From the observations and interviews with staff in the process evaluation, a lack of engagement and poor behaviour spoiled the delivery of some sessions. This may have contributed to the lack of impact in the intervention.

Further analysis suggests that the dialogic reading component of Chatterbooks Plus was possibly too limited to see an effect.

How secure is this finding?

Impact was assessed through a three-arm pupil-randomised controlled trial in twelve schools. 577 Year 7 pupils were randomised either to receive Chatterbooks or Chatterbooks Plus, or to a waitlist control group. Although there was attrition of 21% by the final analysis, there was no evidence that this led to bias or a loss of power that might impinge on the security of the findings.

There had been no formal assessment of the impact of Chatterbooks prior to this trial although some case study work has been carried out in a small number of primary schools. Chatterbooks Plus is a new form of the intervention and has not been trialled in any form. However, there is some evidence from the United States that the dialogic reading element has an impact on younger readers in terms of improvement in oral language. It is important to note that both variants evaluated in this trial were changed significantly from the established Chatterbooks programme that has been running since 2001 as an extracurricular offer.

As the present study represents the first formal evaluation of both programmes and was closely managed by the deliverer, it can be regarded as an efficacy trial. Efficacy trials seek to test evaluations in the best possible conditions to see if they hold promise. They do not indicate the extent to which the intervention will be effective in all schools since the participating schools are selected from one area, and the programme is delivered by the developers.¹

¹ For more details on efficacy trials and other variations of trials, refer to the EEF website: <http://educationendowmentfoundation.org.uk/evaluation/evaluation-glossary/>

The primary outcome was reading ability as assessed by scores from the GL Assessment New Group Reading Test (NGRT). The secondary outcomes were the two NGRT subscales: sentence completion and passage comprehension.

All data was collected by a team from Coventry University who were involved in the Chatterbooks project, but not involved in carrying out the Chatterbooks sessions. Every effort was made to ensure that test administration was blind in every school although complete blindness, as if delivered externally, cannot be guaranteed. The test marking was carried out by GL Assessment and was therefore blind.

Analysis was completed on an ‘intention to treat’ basis, reflecting the reality of how interventions are delivered in practice.²

How much does it cost?

Chatterbooks training costs £1,000 per session for 20 people. A ‘Chatterpack’ for each child costs £5. For a school that trains 20 people and provides the Chatterbooks programme to 150 pupils, the cost will be between £10 and £20 per pupil depending on the amount of supply cover used. The lowest figure in this estimate is based on a non-teaching member of staff delivering the intervention and therefore not requiring cover.

Key conclusions

1. There was no evidence of impact of either Chatterbooks or Chatterbooks Plus on reading ability.
2. The process evaluation indicates that schools encountered problems with timetabling, planning and resourcing, which make the intervention unsuitable in its current format.
3. If implementing this intervention with low-achieving pupils, it would be important to find a way of increasing motivation to mitigate behavioural issues.
4. The dialogic reading component of Chatterbooks Plus could be expanded and trialled separately.
5. To make Chatterbooks more suitable for a school setting, it would be important to consider how sessions that focus on reading for pleasure can be developed to demonstrate impact on reading ability.

² For more details on ‘intention to treat’ and ‘on treatment’ analysis, see the Evaluation Glossary on the EEF website: <http://educationendowmentfoundation.org.uk/evaluation/evaluation-glossary/>

Introduction

Intervention

Chatterbooks is an extracurricular reading initiative designed and delivered by the Reading Agency (<http://readingagency.org.uk>) that aims to increase a child's motivation to read by providing schools with tools and resources to encourage reading for pleasure and to widen the range of genres and authors that they are reading. It consists of weekly small-group sessions intended for six- to twelve-year-olds where attendance is voluntary. In general, the programme is run out of libraries and is flexible in terms of group size. For the purposes of this evaluation, delivery was altered significantly to test its impact in a more structured and formalised school environment. Chatterbooks Plus takes the basic Chatterbooks programme and replaces the last fifteen minutes of a normal session with dialogic reading where children read aloud and are offered explicit prompts.

Chatterbooks Reading Groups aim to encourage reading for pleasure, which the Government promotes as a key part of its commitment to improving literacy skills. For the purposes of this evaluation it is assumed that the enjoyment of reading and discussing books will improve a child's motivation to read independently, resulting in an improvement in reading ability. Consequently, impact is measured as a child's reading ability.

The rationale for this evaluation was to determine any impact on reading attainment both immediately after intervention delivery and at delayed post-test. It was also intended to determine whether, for the programme to be optimally effective in this age group and school context, the approach needed to be more structured and formalised for it to work.

Table 1: The Chatterbooks Logic Model

Intervention	Inputs	Activities	Outputs	Short-term outcomes	Long-term outcomes
Chatterbooks	50 mins a week of time Suitable 'non formal' location Adequate bookstock for children to choose texts from each week Session run by 'tutor' with good knowledge of children's literature and enthusiasm for reading	Discussion of children's books and different authors / genres 'Playful' activities which allow children to explore ideas related to children's literature (eg craft activities, quizzes, role play)	Production of 'lesson plans' for nine Chatterbooks sessions Production of craft and topic work related to children's literature	Increased motivation to engage with books Increased motivation to read for pleasure outside of school	Increased motivation to engage with books Increased motivation to read for pleasure outside of school Increased attainment in reading / English outcomes as a consequence of increased exposure to print and reading activity
Chatterbooks Plus	50 mins a week of time Suitable 'non	Discussion of children's books and different authors / genres	Production of 'lesson plans' for nine Chatterbooks	Increased motivation to engage with books	Increased motivation to engage with books Increased

Intervention	Inputs	Activities	Outputs	Short-term outcomes	Long-term outcomes
	<p>formal' location</p> <p>Adequate bookstock for children to choose texts from each week</p> <p>Session run by 'tutor' with good knowledge of children's literature and enthusiasm for reading</p>	<p>Inclusion of a structured 'dialogic reading' framework for discussing an extract of a book read by the tutor</p> <p>Use of playful activities designed to build the skills known to be associated with reading comprehension</p>	<p>Plus sessions</p> <p>Production of craft and topic work related to children's literature</p>	<p>Increased motivation to read for pleasure outside of school</p>	<p>motivation to read for pleasure outside of school</p> <p>Increased attainment in reading / English outcomes as a consequence of increased exposure to print and reading activity</p> <p>Improved comprehension skills</p>

Background evidence

Previous to this trial there has been no formal assessment of the impact of Chatterbooks on outcomes. Some case study work has been carried out in a handful of primary schools. Currently, there are relatively few secondary schools running Chatterbooks Reading Groups (around 30), but a larger number of primary schools (at least 80). Chatterbooks Plus is a new form of the intervention and has not been trialled in any form. However, the dialogic reading element has evidence of effectiveness in younger readers in terms of improvement in oral language (US Department of Education, Institute of Education Sciences, 2007). As the present study represents the first formal evaluation of the programmes and was closely managed by the deliverer, it can be regarded as an efficacy trial.

Evaluation objectives

The impact evaluation sought to determine the impact of Chatterbooks and Chatterbooks Plus on reading ability. Furthermore, it aimed to discern whether any improvements in attainment were moderated by the National Curriculum reading level of the pupil or whether a pupil received the Pupil Premium.³

The purpose of the process evaluation was to assess Chatterbooks and Chatterbooks Plus in terms of fidelity to the programmes' intentions and their scalability.

Project team

The Chatterbooks programme was designed by the Reading Agency, with the extra elements for Chatterbooks Plus designed by Professor Clare Wood. Both programmes were delivered by graduates at Coventry University hired specifically for this purpose. A further team of graduates, who had no role in intervention delivery, was responsible for test administration. The evaluation team at NFER was led by Dr Ben Styles. Sally Bradshaw and Alix Godfrey assisted with the impact evaluation. Rebecca Clarkson and Katherine Fowler carried out the process evaluation.

³ As data on eligibility for free school meals is more easily obtainable, this measure was used as a proxy for receipt of pupil premium.

Ethical review

Headteachers consented to the trial being carried out within their schools. This consent was followed up by a letter to parents allowing opt-out consent. The pattern of headteacher consent followed by parental opt-out consent, as adopted for other EEF transitions trials run at NFER, was approved by NFER's Code of Practice Committee on 23 January 2013. As a requirement of Coventry University, opt-out pupil consent was also sought for this trial prior to randomisation.

Trial registration

This trial has been registered on <http://www.controlled-trials.com/> number ISRCTN88327135.

Methodology

Design

The project was run as a randomised controlled trial, with 577 individual Year 7 pupils across 12 secondary schools randomised at pupil level to three groups: Chatterbooks and Chatterbooks Plus (intervention groups) and a waitlist control. Pupils in the intervention groups received the appropriate intervention for an hour a week, for nine weeks; pupils in the control group experienced their usual school curriculum. Originally, ten weekly sessions of the interventions had been planned; however, owing to time constraints, in some schools only nine of these sessions took place. Pupils were tested for reading ability before the intervention, immediately after the intervention and then after the summer holidays.

This design was chosen in order to test both whether Chatterbooks itself is effective and whether the addition of a dialogic reading component in Chatterbooks Plus improved effectiveness compared to business as usual.

Eligibility

Year 7 pupils with a National Curriculum Level 4 or below in English and/or reading at the end of Key Stage 2 were invited to take part in this trial. Furthermore, based on pre-existing school data, pupils deemed to be 'vulnerable' Level 4 readers, as indicated by reading ages of 10 or below, were also eligible. Prior to randomisation, opt-out consent was sought from parents of pupils recruited to take part in the trial.

Intervention

Chatterbooks is an extracurricular reading initiative, usually delivered by librarians⁴ trained by the Reading Agency, that aims to increase a child's motivation to read. It consists of weekly small-group sessions, usually in a public or school library, where children read and discuss an age-appropriate, enjoyable book. In this evaluation, nine weekly sessions each of around an hour long were delivered in various locations within school, for example the school library or an empty classroom. The emphasis is on engaging children and encouraging creativity, rather than delivering instruction.

Chatterbooks Plus takes the basic Chatterbooks programme and replaces 15 minutes of each session with a period of dialogic reading in which children read aloud and are offered explicit prompts on relevant vocabulary and situational knowledge in order to enhance comprehension. Dialogic reading is an established approach for much younger readers (US Department of Education, Institute of Education Sciences, 2007), but has not been adapted for or tested on an older group. Dialogic approaches to engaging children in talk around books tend to follow a series of structured prompts – for Chatterbooks Plus, Professor Wood developed the SPICE prompts to elicit talk-based activities that would develop reading skills related to vocabulary, narrative understanding, reading with expression and general knowledge (all elements linked to successful text comprehension skills). The components of a SPICE sequence are:

- **Share a text:** Read aloud a text.
- **Prediction:** Discuss what might happen next.

⁴ In this trial, researchers at Coventry University delivered the programme.
Education Endowment Foundation

- **Improve it:** Identify what you liked most and least, what you would change.
- **Cognitions:** What might characters be thinking?
- **Emotions:** What might characters be feeling?

SPICE was the main structural element, but a choice of follow-up activities were provided that were intended to help the students consolidate comprehension-related skills such as vocabulary, reading expression and visualising the context of a passage of text.

Chatterbooks and Chatterbooks Plus were delivered by trained graduates in 12 schools, over a nine-week period. The eight graduates were also randomised to the two interventions, four to each, to minimise teacher bias. The tutors of the two interventions were encouraged not to discuss their work with each other to minimise cross-contamination between the two interventions. Furthermore, tutors were unaware of which intervention was intended to be superior, the two programmes having been named 'A' and 'B'.

The control group in each school received their normal curriculum.

Outcomes

Raw score from the New Group Reading Test (NGRT; GL Assessment⁵) was used to measure reading ability. At baseline, the digital version of this test was used, while at follow-up the paper version was used due to technical difficulties experienced with the digital version and reservations surrounding the content of the test. The digital test is adaptive. The items that each pupil sees are dependent on whether they have got previous items right or wrong. It thus covers a greater age and ability range than individual paper tests. In particular, if a pupil is struggling with the sentence completion items, the test defaults to phonics items that are not present in the equivalent paper version of the test. When this feature of the test became apparent, concern was raised over whether the follow-up test would be measuring what was originally intended: a reading score made up of sentence completion and passage comprehension items. Along with all other transition trials run by NFER, the decision was made to move to paper testing in order to be confident in the measurement of outcomes at the end of the trial.

Although the use of two different forms of the GL test does not allow a calculation of gain scores, it does not detract from the security of the findings. The role of the baseline in such a model is predominantly to explain outcome variance and hence increase power. When analysing the results of an RCT with a baseline measure, the most unbiased approach is to treat the baseline as a covariate in either ANCOVA or regression (Senn, 2006). As EEF's paper on pretesting⁶ discusses, the measure need not be the same or a parallel test to the post-test. As long as it correlates well with the post-test, it serves its function. In this case, the digital pre-test can almost be regarded as a parallel test to the post-test since it largely contains the same items as the paper form A of the test. The correlation between the digital pre-test and paper post-test was $r=0.64$; thus the pre-test explained approximately 40% of the post-test variance. This was sufficient for the power of the analysis to be retained at the level intended (see 'Minimum detectable effect size' section on page 11).

The NGRT has two parallel forms; form A was used at baseline and form B was used at both first and second follow-up. This was done to ensure a repeated-measures analysis could be carried out after second follow-up thus increasing power.

⁵ <http://www.gl-assessment.co.uk/products/new-group-reading-test>

⁶ http://educationendowmentfoundation.org.uk/uploads/pdf/Pre-testing_paper.pdf

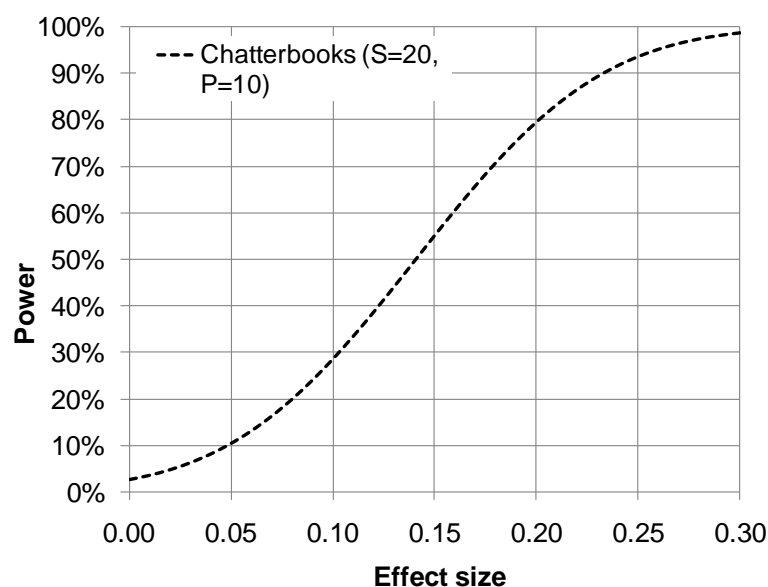
The NGRT has two subscales: sentence completion and passage comprehension, which can be combined into a composite reading score. The composite score was used as the primary outcome. The two subscales were used as secondary outcomes. These outcomes were chosen since one aim of the intervention is to improve reading ability and the NGRT is a reliable test of reading that has been standardised for the age group in question.

All data was collected by a team from Coventry University who were involved in the Chatterbooks project, but were separate from the graduates employed to lead the sessions. At baseline, this team invigilated while pupils took the digital tests. The marking was calculated blind by GL Assessment's online system and results were accessed through GL Assessment's online platform following testing. The same team invigilated both phases of the paper testing, and sent the completed scripts to GL Assessment for blind marking.

While the test administrator team was not involved in carrying out the Chatterbooks sessions, at least one member was aware of the results of the randomisation, and had observed some intervention sessions, so may well have been aware of the condition of some pupils. We therefore cannot claim that administration was blind in every school, as it would have been if administered externally, but attempts were made to ensure it was as blind as possible.

Sample size

Figure 1: Power curve



S = Number of participating schools
P = Number of participating pupils per intervention group per school

Randomisation was conducted at the pupil level, and variation in baseline scores was controlled for in the final analysis. Intra-class correlation (ρ) was therefore likely to have a minimal impact on the effective sample size, and so a low value of $\rho=0.02$ was assumed for the purposes of sample size calculations. We also assumed a correlation of 0.75 between baseline and follow-up scores on the basis of previous work with reading tests. Figure 1 illustrates that a sample size of 200 per intervention group should be sufficient to detect effect sizes of the order 0.20. This could be considered low–moderate, equivalent to around three months of progress – quite reasonable for targeted interventions providing support to small groups of pupils.⁷

It proved difficult to recruit 20 schools and priority was given to recruiting enough pupils to the trial. As this was a pupil-randomised trial, the number of schools was not so crucial and there were still sufficient numbers of eligible pupils in the 12 recruited schools to carry out an adequately powered trial.

Minimum detectable effect size (MDES)

Once all the data from the trial was available, the assumed parameters from the above calculations were compared to the actual parameters and included in a calculation of MDES.

Randomisation was carried out at the pupil level thus cancelling out the effect of clustering when estimating internally valid uncertainty around the effect. ρ can hence be regarded as zero. The adjusted R-squared for the primary outcome model without the intervention terms was 0.46, implying a value of 0.68 would have been more appropriate for the correlation between baseline and follow-up scores. Using the actual number randomised, this yields an MDES of 0.22 at 80% power.

⁷ Effect sizes are for paired comparisons between two of the three groups (eg Chatterbooks vs control or Chatterbooks Plus vs Chatterbooks). These differential effects will be smaller, and so are less likely to be detected for a given sample size.

Randomisation

Researchers at Coventry University and the individual schools involved in the trial were responsible for pupil recruitment. Randomisation was carried out by a statistician at NFER using a full syntax audit trail within SPSS. Randomisation was stratified by school: simple randomisation of pupils into three groups of the same size was carried out within each school. This was necessary to aid timetabling of the sessions within schools.

Simple randomisation of tutors between Chatterbooks and Chatterbooks Plus groups was also performed. Furthermore, in schools with a large enough group of eligible pupils, intervention pupils underwent further random allocation to teaching groups.

Analysis

The primary outcome was reading ability as assessed by the NGRT. Sub-group analysis on the primary outcome was carried out on the following groups only: pre-test score, National Curriculum level and whether a pupil was eligible for free school meals (FSM). The secondary outcomes were the two NGRT subscales: sentence completion and passage comprehension. All outcomes and sub-group analyses were pre-specified at protocol stage.

The definitive analysis was 'intention to treat', reflecting the reality of how interventions are delivered in practice. The main analysis of the complete dataset consisted of repeated-measures models retaining pre-test score as a covariate. These models were run in MLwiN (three levels: time, pupil and school) and included interaction terms between time and intervention to discern whether changes over time were significantly different between the three experimental groups. It was necessary to take school into account in the analysis due to the restricted nature of the randomisation (Kahan and Morris, 2012). For the repeated-measures model, this was done by including school as a level. For the calculation of effect size, a series of dummy variables were included in a simple regression model. The definitive primary outcome analysis regressed post-test raw scores in a repeated-measures model on pre-test score, randomised group, sex, age in months, FSM, time and an interaction between randomised group and time. Subgroup analysis was carried out by exploring the interaction between randomised group and pre-test score and this was followed up by analysis by National Curriculum level. Further subgroup analysis explored the interaction between randomised group and FSM. Secondary outcomes were analysed using raw scores in the relevant domains in place of overall reading scores.

The main analysis was followed by an 'on-treatment' analysis where data from the tutor logs was used to determine the extent of each pupil's involvement with the interventions. This analysis allows for an estimate of a 'pure intervention effect' (net of any fidelity issues, contamination or non-completion). However, note that this analysis may be biased due to self-selection to differing levels of exposure.⁸

Process evaluation methodology

The process evaluation covered the entire length of the intervention from start-up in March 2013 to completion in July 2013.

Information was collected from the following sources: a review of the training materials and observations of the training sessions, observations of intervention sessions in situ, telephone interviews with the deliverers, and reviewing the qualitative parts of the 'tutor logs'. Interviewees were

⁸ For example, pupil motivation may be positively related both to levels of exposure to the intervention (through better attendance) and to the amount of progress made between baseline and follow-up testing.

selected at random from those that had not already been observed delivering sessions. Further information was also gathered in informal discussions with the delivery partner. These methods were chosen to allow for coverage across different parts of the intervention.

A team of researchers from NFER conducted the process evaluation. Where more than one researcher was observing training and sessions and undertaking interviews, frequent team meetings were held to share information and plan next steps. All researchers have contributed to the report writing to ensure full coverage of information gathered.

Detailed training and sessions observation schedules were developed to ensure that comparative information was gathered. Part of the 'tutor log' was also used to capture information about what was happening in each of the sessions. This included a space to record a brief outline of each session and a section to detail whether any deviation from the programme had occurred.

Four individuals from the pool of deliverers were interviewed by telephone, two who delivered Chatterbooks and two who delivered Chatterbooks Plus. None of these deliverers had been previously observed as part of the session observations. An interview schedule was developed to ensure consistency in questioning so that comparable information could be gathered. Interviews took approximately one hour to complete.

The telephone interviews covered questions and topics on the training, delivering the interventions, resources needed and used, issues of cross-contamination, and other issues such as perceptions of scalability. The responses have been used to inform the detail in the sections describing the training and intervention sessions in schools.

Impact evaluation

Timeline

Table 2: Timeline

Month (2013)	Activity
January – February	Recruit and gain consent from schools and pupils
March	Recruit and consent schools and pupils Random allocation of pupils ⁹ Pre-testing (18–28 March) Training of delivery researchers
April	Implementation of intervention programmes
May	Implementation of intervention programmes
June	Implementation of intervention programmes Post-testing (1 st follow-up; 24 June – 12 July)
September	Post-testing (2 nd follow up; 23 September – 4 October)
October onwards	Waitlist control pupils receive Chatterbooks Plus

Participants

Schools involved in this project were located across the Midlands, for example Birmingham, Solihull and Warwickshire. Further details are provided in Tables 2–5. Schools were recruited to the project by Professor Wood from Coventry University, through emails to schools in the nominated areas and follow-up face-to-face meetings with those that showed initial interest. Schools were required to sign a memorandum of understanding (see Appendix). Recruitment, and headteacher and parental consent, took place between January and March 2013. A detailed timeline including recruitment, testing and intervention implementation can be found in Table 2.

Across 12 schools, 2524 pupils were checked for eligibility to the study by staff in the individual schools, using the criteria provided by EEF and Coventry University. The majority of schools discussed this eligibility assessment with the team at Coventry University, in order to ensure the appropriate pupils were entering the study. Of the pupils assessed, 577 were deemed eligible and were randomised to the intervention or control groups.

Table 3: School setting

School setting	Number of schools
Urban	11
Rural	1

⁹ Baseline testing was scheduled to finish on 28 March but, for timetabling reasons, some randomisation results had to be released to test administrators on 25 March so there was overlap. This was necessary owing to the constrained timescale and schools' requirement to know groupings early for timetabling reasons.

Table 4: Ofsted rating for schools

Ofsted rating of overall effectiveness of the school	Number of schools
Outstanding	3
Good	5
Satisfactory	3
Inadequate	1

Table 5: School type

School type	Number of schools
Academy	9
Comprehensive 11–16	1
Comprehensive 11–18	2

Table 6: Pupils eligible for FSM

Pupils eligible for FSM	Number of schools
Highest quintile	3
Second highest quintile	3
Middle quintile	2
Second lowest quintile	2
Lowest quintile	2

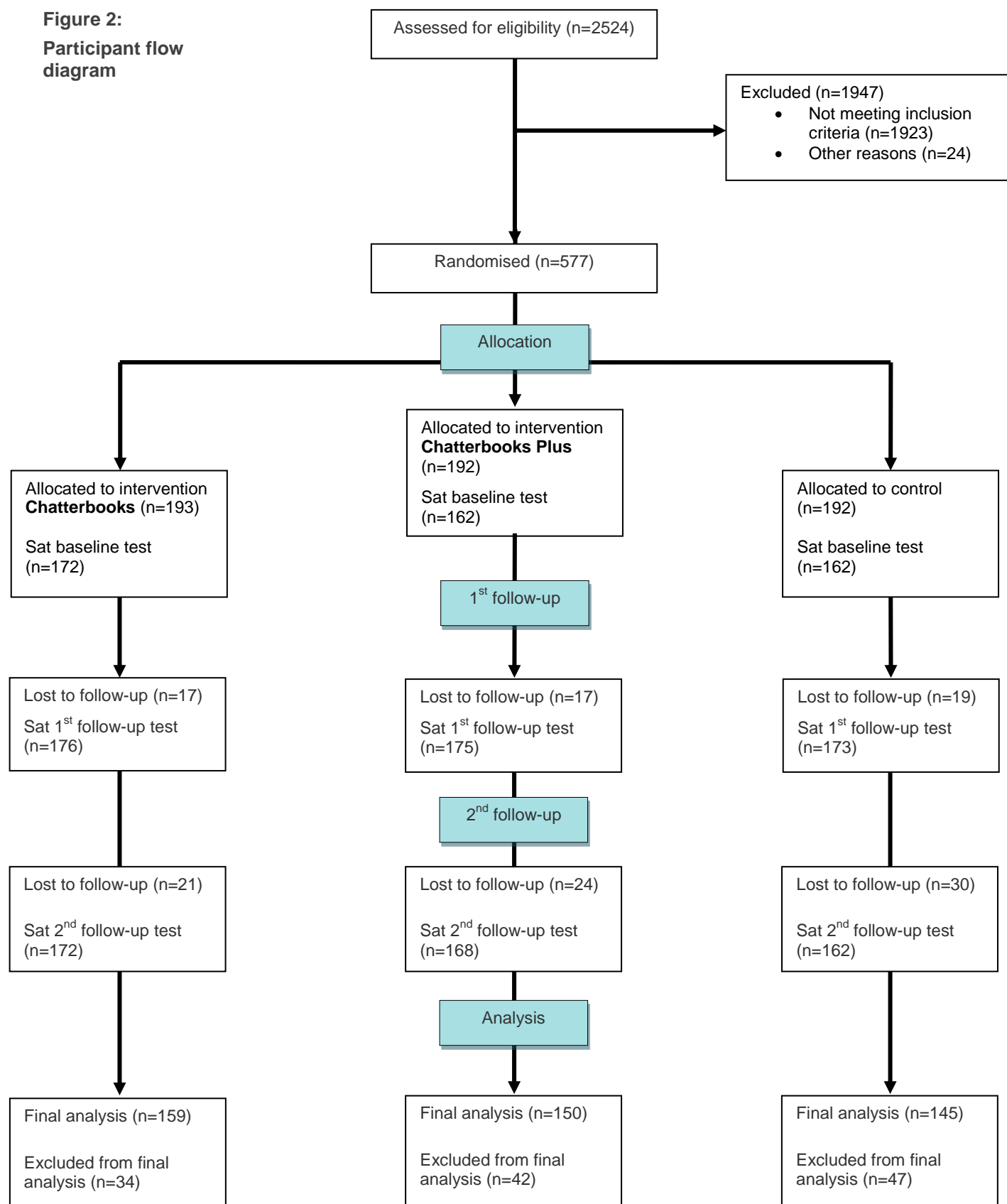
Table 7: School attainment

GCSE performance band (2012)	Number of schools
Highest quintile	4
Second highest quintile	5
Middle quintile	2
Unknown	1

All schools involved were secondary, mixed-sex schools in the West Midlands, with Ofsted ratings ranging from *Outstanding* to *Inadequate* at their last inspection. With one exception, all schools in the sample were in urban areas.

The number of ethnicity categories per school ranged from 13 to 17 (out of a maximum of 18 Office for National Statistics categories), suggesting that all schools had ethnically diverse pupils. The sample also showed diversity in terms of deprivation, with two schools in the bottom quintile for FSM eligibility, and three in the top quintile. School attainment information was available for 11 of the 12 participating schools. Of these, all were at or above the middle quintile, based on their GCSE performance band.

Figure 2:
Participant flow diagram



Based on information provided by the delivery partner, four pupils either left their schools, or were excluded from school, during the project. These pupils were removed from the trial. The remaining pupils were put forward for baseline and follow up testing regardless of their cooperation with the interventions.

The vast majority of screened pupils who were excluded from the project were not eligible based on the inclusion criteria. There were several reasons for the exclusion of the remaining pupils: they declined to take part, their parents refused consent, or they were excluded from school in between screening and randomisation.

Pupils who did not sit both pre- and post-tests were excluded from the final analysis, owing to inability to compare their results at the two time periods.

While Figure 2 suggests randomisation occurring before baseline, in reality there was a degree of crossover, as some schools required results before their pupils sat the baseline test, for timetabling reasons.

Pupil characteristics of analysed groups

While we expect no systematic bias to have arisen from randomisation, bias may have occurred owing to attrition. Chi-squared tests on all three background factors presented in this section revealed no significant differences between groups for the data after attrition (Tables 8–10).

Table 8: National Curriculum level in reading at baseline ($\chi^2 = 10.1$, $df=8$, $p=0.26$)

National Curriculum level (source: baseline NGRT)	Chatterbooks		Chatterbooks Plus		Control	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
1 or below	19	12	21	14	18	12
2	53	33	45	30	59	41
3	63	40	59	39	39	27
4	23	15	23	15	25	17
5	1	1	2	1	4	3
Missing	0	0	0	0	0	0
Total	159	100	150	100	145	100

Table 8 shows that the vast majority of pupils satisfied the eligibility criteria as applied by schools. It should be noted that schools were using Key Stage 2 results to determine eligibility criteria so some improvement in reading level may have occurred since then.

Table 9: FSM eligibility ($\chi^2 = 2.2$, $df=2$, $p=0.33$)

Pupil eligible for FSM (source: schools via GL Assessment)	Chatterbooks		Chatterbooks Plus		Control	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Yes	37	23	40	27	44	30
No	113	71	104	69	91	63
Missing	9	6	6	4	10	7
Total	159	100	150	100	145	100

Table 10: Sex ($\chi^2 = 1.3$, $df=2$, $p=0.52$)

Sex of pupil (source: schools via GL Assessment)	Chatterbooks		Chatterbooks Plus		Control	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Male	82	52	87	58	80	55
Female	77	48	63	42	65	45
Missing	0	0	0	0	0	0
Total	159	100	150	100	145	100

Further to pupil background measures, it was also important to test whether significant imbalance at pre-test had ensued as a result of attrition. The baseline effect size was 0.00 (-0.23, 0.24) for Chatterbooks and -0.01 (-0.23, 0.21) for Chatterbooks Plus. There was no significant imbalance at baseline ($p=0.99$).

Outcomes and analysis

Table 11: Effect sizes at second follow-up

Outcome description	Outcome measure	Chatterbooks vs control effect size (Hedges' g and 95% confidence interval)		Chatterbooks Plus vs control effect size (Hedges' g and 95% confidence interval)		Number of Chatterbooks pupils in model	Number of Chatterbooks Plus pupils in model	Number of control pupils in model
Primary	Reading score	-0.14	(-0.31, 0.03)	-0.01	(-0.18, 0.16)	159	150	145
Primary (FSM pupils)	Reading score	-0.30	(-0.63, 0.03)	-0.09	(-0.39, 0.21)	37	40	44
Secondary	Sentence completion score	-0.18	(-0.37, 0.01)	0.00	(-0.19, 0.20)	159	150	145
Secondary	Passage comprehension score	-0.10	(-0.28, 0.09)	-0.03	(-0.22, 0.16)	157	148	144

An ANOVA of first follow-up reading score by randomised group revealed no significant impact ($F=0.59$, $p=0.56$, $n=524$). An ANOVA of second follow-up reading score by randomised group had similar results ($F=0.50$, $p=0.60$, $n=502$).

The main analysis of the complete dataset consisted of repeated-measures models retaining pre-test score as a covariate. This model was run in MLwiN (three levels: time, pupil and school) and included interaction terms between time and intervention to discern whether changes over time were significantly different between the three experimental groups. None of sex, FSM eligibility and age in months was significant in the model and these variables were therefore excluded. A separate model was run on FSM pupils and interactions with pre-test score were explored in the main model. Secondary outcomes were also modelled in this way. As expected, time of measurement was significant: post-test score had increased by 0.98 (0.49, 1.48) points between first and second follow-up. However, no significant effects of the interventions were found across either follow-up. Note that for this model, data had to be present from pre-test and at least one of the follow-up tests to be used.

All outcomes analysed were pre-specified in the protocol. All sub-group analyses were pre-specified in the protocol apart from the use of FSM as a proxy for pupil premium; a separate FSM analysis is a requirement of all EEF projects. Background data on pupils was obtained both from schools through the standard GL Assessment data form and from the National Pupil Database (NPD). Where the same variable was obtained from both sources, fewer missing cases were seen using data collected from

schools due to the imperfect match to NPD using UPN. Schools' data was hence used in preference where there was a choice.

Effect sizes were calculated on second follow-up data, three months after the first follow-up test, since this was viewed as more important than first follow-up as it demonstrates a degree of retention of any gains in reading ability. Effect size calculations used a regression model containing the following variables: pre-test score, intervention group, school and age in months. FSM and sex were included in an initial run of the model but neither was significant so both were excluded. Effect sizes were calculated like this rather than from the repeated-measures multi-level model as this is more consistent with EEF reports in general.

The protocol specifies both an analysis by National Curriculum level and pre-test score. This was also explored using repeated-measures models. There was no significant interaction between intervention group and pre-test score. As expected, since the analysis by pre-test score is more sensitive, an analysis by National Curriculum level revealed no significant effect of either intervention.

All the above analysis was 'intention to treat'. An 'on-treatment' analysis was carried out using data collected and entered by the Coventry University team. The measure used was a count of the number of sessions each child experienced; Chatterbooks and Chatterbooks Plus were considered equally for this analysis. For Chatterbooks, the median number of sessions an individual experienced was nine and 62 (34%) pupils received ten sessions. For Chatterbooks Plus, the median number of sessions an individual experienced was also nine and 58 (33%) received ten sessions. Control pupils were considered to have experienced no sessions since there was no evidence of contamination. As for the definitive analysis described above, a repeated-measures model was used for the dosage analysis. It revealed no significant effect: having one more session was associated with a change in post-test score of -0.02 (-0.12, 0.08).

Cost

Currently, Chatterbooks training is £1,000 for a session, which can be delivered to around 20 people. There is also a £5 charge for the 'Chatterpacks' which each child receives. For a school that trains 20 people and provides the Chatterbooks programme to 150 pupils, the cost will be between £10 and £20 per pupil depending on the amount of supply cover used. The lowest figure in this estimate is based on a non-teaching member of staff delivering the intervention and therefore not requiring cover. However, the Reading Agency is rethinking its model for costs of materials.

Professor Wood and her team produced lesson plans for the purposes of this intervention. If schools decided to deliver Chatterbooks this is something that the deliverer would need to do in their own time. The Reading Agency also has materials on their website. They suggested that staff planning time would be minimal, but this was not backed up by comments received from the telephone interviews (see 'Process evaluation' on page 20). Schools do not need to purchase any specific ICT equipment. The team at Coventry University supplied plasticine at a cost of 99p per child; otherwise children used pens and paper, which are readily available in schools. The Coventry team used digital voice recorders in one session so that children could 'write' orally – these would cost £20–50.

The schools involved in the trial did not incur any costs. All costs were covered by the team at Coventry University; this included an allowance to enable the schools to refresh the book stock in the school library to optimise the book choices that the children could make. Schools paid the supply cover for the teachers to attend the Chatterbooks training. This would be roughly £180 per teacher.

Process evaluation

The two strands of the intervention, Chatterbooks and Chatterbooks Plus, were sketched out and finalised before training took place. Recruitment of the delivery team for each strand was also completed before training. The training was delivered by the Reading Agency, which has a history of training and delivering Chatterbooks. The extra element for Chatterbooks Plus, 'SPICE', was devised by Professor Wood and training for this was delivered on the same day as the rest of the training.

The deliverers were randomised into two groups, with one group receiving training and delivering sessions just in Chatterbooks, and the other group concentrating solely on Chatterbooks Plus. The two strands of the intervention were delivered in each school, so deliverers were instructed not to mix between groups.

Training observations

Each training day was specific to the different strands of Chatterbooks – the deliverers delivering Chatterbooks Plus were not present at the Chatterbooks training day and vice versa. Training consisted of a one-day course held at Coventry University in March 2013 and was run primarily by the Reading Agency and a representative from the Schools Library Service for Coventry. Professor Wood arrived towards the end of the day to conduct a Q&A session with delegates where she was able to clarify points from the morning's research section of the training. All delegates were given a training pack on arrival, which consisted of the day's programme, Chatterbooks ideas sheets, a Chatterbooks skills checklist (necessary skills for running Chatterbooks), a sheet of basic tips and a planning sheet. For Chatterbooks Plus the training pack also included a sheet detailing the required activities and structure of each session.

Training for Chatterbooks was attended by 20 delegates made up of research assistant deliverers and teachers from the schools participating in the project, with the majority being teachers. Teachers were invited as an incentive for schools to take part in the project. For Chatterbooks Plus, there were four delegates only – the deliverers; no teaching staff were invited as Chatterbooks Plus is not endorsed by the Reading Agency. The delegates were not made aware in the training that this was not the original format of Chatterbooks.

The concept of Chatterbooks was introduced, including several case studies of children in current groups around the country whose enthusiasm for reading has increased owing to their involvement in the scheme. There was also discussion about the desired outcomes of Chatterbooks, that these would be measured using an unspecified reading test, and how to achieve them. In the afternoon, the training focused on the outline of a single Chatterbooks session. The majority of the afternoon was given over to planning, with delegates working in groups to create session plans.

There was some confusion about how schools were going to access the books required for the sessions. There were problems with engagement from the teaching staff; as they were not delivering, the training did not seem beneficial.

Session observations and tutor logs

Chatterbooks

Two Chatterbooks sessions were observed, each lasting approximately one hour. Both of these took place on the same morning (18 June 2013), at a secondary school in Birmingham. Two different deliverers were observed. The theme of the week was 'film week', the eighth session out of nine planned, and both sessions had the same overall session plan.

Pupils were extracted from a variety of other lessons, often ones which they enjoyed, for example history. The sessions took place in the large school library, with assistance from an enthusiastic

school librarian, who especially sought out the pupils to attend. It is therefore not surprising that attendance at Chatterbooks sessions at this school was high with only one or two pupils absent.

There was plenty of space and furniture for all participants to be comfortable. The deliverers used two whiteboards in the sessions. During the first session, another Chatterbooks group was being run in parallel at the other end of the library. Due to the library's large size, this did not affect the group.

Both sessions began with an icebreaker activity, followed by a quiz-show style activity. The sessions ended with the pupils being given an opportunity to choose from a selection of books that fitted in with the week's theme. They were able to take these books home with them if they so wished.

Chatterbooks Plus

One Chatterbooks Plus session was observed, also during 'film week', on 14 June 2013 at a different secondary school in Birmingham. This session was also in the morning and one hour long. There were four pupils who had been extracted from various lessons; attendance was low owing to timetable changes at the school. This meant there was plenty of space and seating arrangements were adequate.

The session started with an icebreaker activity. Pupils then watched the opening scene of a film and did the SPICE activity based on this scene; however, there was no evidence of the *Cognitions* or *Emotions* components of SPICE. The session deliverer then read out the opening passage of the book which corresponded to the film scene. They had a brief discussion about the differences between the film and the book. They then watched a nine-minute clip from another film, which was the starting point for another activity. The session finished with a very brief discussion about what the children had been reading since the last session.

All the observed sessions were at the same point in the schedule, suggesting that the intervention programme was being adhered to faithfully.

Implementation

A key aspect of the success of Chatterbooks and Chatterbooks Plus, based on discussions and observations, appears to be the motivation of students. If a wider roll-out were justified, it would be necessary to find some way to keep students motivated to turn up to sessions. It was clear that this was dependent on the level of proactivity of the school. Schools' engagement would be crucial in achieving a wider roll-out of the programme.

When asked about the main barriers to implementing Chatterbooks and Chatterbooks Plus, a common theme mentioned by deliverers was behavioural problems within the sessions. Time spent managing behaviour meant that on occasions not all of the required activities could be delivered. Some students missed out on some activities, or did not get the full benefit of them. Deliverers suggested that in some schools Chatterbooks was not taken seriously and seen by some as a 'doss session', in the words of one deliverer. Or, similarly, that the children meant to be taking part saw this as an undesirable thing as it took them away from other, seemingly more interesting, school activities. This latter view was based on feedback from pupils at a session where one pupil had decided to stay in their usual lesson and was backed up by a deliverer of a different session.

The training implied flexibility as a Chatterbooks leader but for the evaluation all the deliverers had to carry out the same programme so that they were comparable. For Chatterbooks Plus, sessions had to be very tightly structured and this was not explained fully in the training. The Chatterbooks Plus deliverers would have preferred more flexibility, but given that there needed to be a strict structure, they wanted this explained.

Fidelity

Structure

There was an intensive level of quality control of the sessions, in terms of adherence to both the research aims and the ethos of Chatterbooks. Session plans, which the deliverers developed as a group (Chatterbooks or Chatterbooks Plus) were scrutinised by both Professor Wood and a representative from the Reading Agency. This meant that, based on the session observations carried out, the sessions seemed broadly to adhere to what was advised in the training.

However, this level of standardisation, for the purpose of the research project, was felt to be at odds with the flexibility inherent in the standard Chatterbooks programme, delivered in the 'real' world. This led to a feeling after the training that not enough information had been given, as deliverers were being expected to deliver something far more structured than Chatterbooks sets out to be. As all Chatterbooks/Plus deliverers were expected to plan their weekly session together, and in theory deliver the exact same session across different pupil groups in different schools, the core flexibility of Chatterbooks, which could be seen as the key to its success, was lost.

With regards to Chatterbooks sessions, the training made it very clear that flexibility and lack of structure were key concepts. Whereas Chatterbooks had only a couple of required elements (the icebreaker and the theme for each week) Chatterbooks Plus sessions were intended to be much more structured with a required set of activities.

The training day was lacking in detail on this structure. Interviews revealed that deliverers did not find planning sessions a straightforward process. There was some debate between deliverers and Professor Wood over the content of lesson plans.

This suggests that if a strict format is to be used for sessions, it should be clear from the start and emphasised in training. For example, we found that deliverers were unsure how much they should encourage children to take books home and read outside of the sessions. It was hard to fit this in given the limited time available and it was not made clear, in the training or subsequently, whether this was a key aim of the intervention.

Another issue was timing. Deliverers were not always able to fit in all the activities on the session plan. This was particularly the case with Chatterbooks Plus, where many activities were expected to be completed in one hour. Deliverers interviewed felt that there were too many activities.

In some schools, only nine of the intended ten sessions were delivered as they were unable to accommodate the tenth session due to exams and other commitments. Sessions nine and ten were often combined, meaning that a large number of activities needed to be fitted into one hour – described by one deliverer as '*chaos*'.

Attendance

When asked about perceptions of average attendance rate across their sessions, all interviewees were satisfied with the level of attendance at their sessions; however, this seemed to be variable across schools. The Chatterbooks session observed had high attendance, whereas the Chatterbooks Plus session observed was poorly attended due to scheduling issues, and was typically poorly attended at that particular school. Some interviewees mentioned that poor attendance could be related to poor communication with schools regarding scheduling.

Outcomes

When asked about the effectiveness of the programme all deliverers interviewed were able to give both positive and negative anecdotal evidence. Positive anecdotes included pupils reading more, talking about books more, and wanting to take books away with them. This final point was also mentioned by the deliverers observed during session observations. Negative anecdotes tended to centre on the pupils with the lowest levels of motivation to read. For these pupils, based on the anecdotes given, Chatterbooks seemed to have little effect on their desire to read.

As for improving reading attainment, deliverers were less sure that their sessions could achieve this. They felt that the actual reading content of sessions was low; the main focus being on keeping children engaged through fun activities, and, on many occasions, managing behaviour.

No unintended consequences or negative effects of the intervention were mentioned by deliverers' interviews, nor were they observed in the sessions.

Scalability

While interviewees felt that Chatterbooks would generally be popular with schools and would be successful in a wider roll-out, there are some practical issues to mention.

Resources

The sessions tended to be quite resource intensive; Chatterbooks Plus deliverers mentioned making lots of resources and games by hand and doing lots of creative activities involving craft materials. For the research project, deliverers were provided with the resources they required, although this differed between the two versions of the intervention. For Chatterbooks deliverers there seemed to be an almost limitless choice of resources, with no mention of a budget, whereas Chatterbooks Plus found resources not to be forthcoming and paid for some themselves. A wider roll-out would, clearly, not come with the same budget for any resource that could be needed, which could make it less attractive to schools.

Time

The other key issue in terms of the potential popularity of Chatterbooks and/or Chatterbooks Plus is the preparation time needed. In all interviews, the deliverers, who worked full time on the project, emphasised the huge amount of time that it took in order to prepare for a week's worth of sessions – something that teachers would not have.

Proactivity

Implementation would require enthusiasm and organisation from schools in order to make the programme work without the assistance of an external agency. As quality of session planning may not be monitored in the same way, delivery in schools could be much more variable.

Teachers

Passion for books and reading as a key requirement for delivering Chatterbooks/Plus was evident throughout the process evaluation activities. If Chatterbooks/Plus was rolled out more widely in schools, there is no guarantee that teachers (who would be delivering it) would value it in the same way; however, delivery by teachers could be positive in terms of behaviour management. In this trial, research assistants delivered the intervention.

Control group activity

Interviewees were probed about whether there had been any contamination between the two versions of the intervention and/or the control group. The deliverers interviewed differed in their opinions about whether there had been any cross-contamination between the two research groups. The view of one deliverer was that they could not be certain about cross-contamination within schools, owing to the short amount of time spent visiting the schools per week. The other was “100 per cent” confident that pupils and school staff were not sharing ideas. There is hence a lack of evidence either way as to whether contamination was a problem.

Interviewees thought that children receiving each of the two versions of the intervention may have discussed what they did in their respective sessions; however, this level of informal discussion is unlikely to have led to cross-contamination, particularly as deliverers were highly aware of the need to keep the sessions to plan.

Deliverers from Chatterbooks became aware quite quickly that the deliverers working on the Chatterbooks Plus arm of the research project were doing something slightly different, and that consequently they were not meant to be discussing the project with them. This suggests a lack of cross-contamination between the two groups of researchers.

Recommendations for improvements

Address behaviour

During the session observations, it was very clear that poor behaviour was common among the students taking part. The deliverers had not been given any form of behaviour management training. If an external agency were to be used, they would need thorough training in behaviour management. Alternatively, it is likely that behavioural issues would not cause such a disturbance if the sessions were led by members of school staff.

Group pupils based on interests

Group dynamics were key to the success of sessions. Deliverers suggested the possibility of grouping students by their interests, in order to make the sessions easier to tailor.

Target appropriate age group

Other comments questioned the appropriateness of Chatterbooks for Year 7s. In practice, Chatterbooks is aimed at primary-age children. One deliverer suggested that the resources provided were more suitable for younger pupils, so these may need adapting. However, it seems that the activities themselves would be adaptable to different age groups.

Chatterbooks Plus – fewer activities

Observation and interviews suggested that Chatterbooks Plus, as intended, consisted of too many activities for the time available – deliverers struggled to fit everything in. For wider roll-out, fewer activities might make for better sessions.

Extracurricular

A suggestion from one deliverer was that Chatterbooks would work better as an extracurricular programme. While this suggestion is most likely true, as this is how Chatterbooks is run in the community, it would also come with the same trappings of the standard Chatterbooks programme, ie it is highly likely that only pupils who already have good levels of literacy, and an interest in reading,

would attend. In order to bring Chatterbooks to pupils who need the increased motivation, it seems likely that it would have to be enforced as part of the timetable.

Conclusion

Limitations

79% of pupils that were randomised were included in the final second follow-up effect size calculations. This was measurement attrition rather than lack of intervention fidelity; all pupils that were randomised and did not subsequently formally withdraw had a post-test administered if they were present at school. For an individual testing sweep, the level of drop-out did not exceed 14% but this final analysis required pupils to have sat both baseline and second follow-up tests. A completers analysis was performed and there is a possibility this was open to bias due to the extent of missing data. However, multiple imputation and sensitivity analyses (Torgerson and Torgerson, 2013) were beyond the scope of this evaluation. In an attempt to reduce attrition, a further analysis was carried out replacing pre-test score with raw score in Key Stage 2 reading. Unfortunately, owing to the imperfect match with NPD this resulted in fewer cases being used in the analysis. Furthermore, the correlation between Key Stage 2 reading and post-test was only $r=0.50$ as compared to $r=0.64$ for the pre-test to post-test correlation.

While internally valid, this trial has little external validity from a strict statistical perspective as the schools were approached non-randomly by Coventry University. This was agreed in the evaluation design and, while very effective as a recruitment strategy, the sample cannot be said to be representative of any meaningful population of schools. However, there is no reason to doubt the similarity of other low-attaining Year 7 pupils sharing similar characteristics to those tabulated in the 'Pupil characteristics' section of this report (see page 17). It therefore seems unreasonable to declare the results not generalisable.

While there was no evidence of contamination, it is possible that pupils might have discussed Chatterbooks and Chatterbooks Plus sessions with their friends and thus perhaps transferred any new enthusiasm for reading onto their friends who might have been in the control group. This is the kind of contamination risk apparent in most pupil-randomised trials and, since the interventions themselves were not particularly amenable to this kind of transfer, contamination is unlikely to represent a risk to the validity of the result. Furthermore, while we cannot be sure tests were administered completely blind, administrators had no role in intervention delivery.

As control pupils were not removed from lessons, it is possible that any improvements in reading seen as a result of the intervention were offset by improvements that control pupils may have achieved by staying in lessons.

Interpretation

Whether a more sophisticated treatment of missing data would have changed the results is doubtful since the baseline characteristics of analysed groups did not differ significantly. This is indicative of an unbiased attrition. Furthermore, it seems unlikely that a trial run in a different area of the country would reveal anything strikingly different in terms of impact since there is no region-specific aspect to the intervention. The limitations above should therefore be seen in the context of fairly conclusive impact results.

The process evaluation highlighted a number of issues surrounding attendance at sessions, behaviour in sessions and motivation of the pupils. This was, perhaps, compounded by the pupils being taken out of lessons that they enjoyed. If the programme were to be implemented, a school would need to think carefully about how to fit these sessions into the school timetable. Pupils would need to be extracted from lessons, the sessions would need to be held over a lunch period, or the sessions would need to be offered as extracurricular. None of these seems suitable, bearing in mind the lack of motivation of the pupils observed. Both researchers who observed sessions were in agreement that there was a large difference in the level of respect that pupils had for permanent school staff as

compared to the researchers delivering the programme. This resulted in behaviour problems that may have seriously impacted on the effectiveness of both interventions if prevalent across all sessions.

Chatterbooks Plus consisted of an extra component of dialogic reading. However, this component was small, lasting typically ten minutes per session, and in the session observed was not fully explored. Given the existing evidence for dialogic reading, the restricted nature of the dialogic component may explain the lack of impact.

There is also a mismatch between the highly structured format of the interventions delivered for this trial and the ethos of Chatterbooks and how 'real' world sessions are delivered. It was evident from the training that Chatterbooks is fairly free-form, with librarians delivering what they wish around a central theme. This is in stark contrast to the weekly meeting and sharing of ideas, auditing of lesson plans and sharing of resources experienced by the intervention deliverers for this trial.

Additional problems for schools who would wish to implement this programme are: the need for the deliverer to be both committed to the intervention and informed in a wider sense about books and reading; and the potential number of resources needed in delivering the sessions, although this would be entirely dependent on the content of sessions.

The impact results indicate that neither Chatterbooks nor Chatterbooks Plus, as delivered in schools for this evaluation, is effective in improving reading skills for low ability Year 7 pupils. Furthermore, impact analysis of the number of intervention sessions experienced by each pupil showed no significant effect, implying the lack of impact was not due to incomplete intervention delivery. Since the dialogic reading component in Chatterbooks Plus was limited and not necessarily implemented as intended, the effectiveness of this element cannot be discounted from these results alone.

Future research

Given the existing evidence base for the dialogic reading component of Chatterbooks Plus (US Department of Education, Institute of Education Sciences, 2007), it seems premature to discount this on the basis of the lack of impact demonstrated in this study. One avenue for further research might be to evaluate an intervention that has this element expanded and trialled separately.

References

Kahan, B.C. and Morris, T.P. (2012). Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis, *BMJ* 2012, 345 (DOI: <http://dx.doi.org/10.1136/bmj.e5840>) [online]. Available: <http://www.bmj.com/content/345/bmj.e5840> [20 December 2013].

Senn, S. (2006) Change from baseline and analysis of covariance revisited, *Statistics in Medicine*, 25, 4334–4344

Torgerson, C. and Torgerson, D. (2013). *Randomised Trials in Education: an Introductory Handbook* [online]. Available: http://educationendowmentfoundation.org.uk/uploads/pdf/Randomised_trials_in_education-revised250713.pdf [20 December 2013].

US Department of Education, Institute of Education Sciences (2007). *Early Childhood Education: Dialogic Reading* (WWC Intervention Report) [online]. Available: http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/WWC_Dialogic_Reading_020807.pdf [20 December 2013].

Acknowledgements

The NFER evaluation team is hugely grateful to Clare Wood, Emma Jackson and their team from Coventry University who recruited schools, managed communication with schools and administered the tests. At NFER, the statistical analysis was undertaken with the help of Sally Bradshaw and Alix Godfrey. Simon Rutt commented on a final draft of the report. We are grateful for the comments and additions provided by Camilla Nevill and James Richardson at EEF, two anonymous referees and Dr Long.

Appendix A: Model Results

Results of main effect model:

Coefficients ^a					
Model	Unstandardised Coefficients		Standardised Coefficients	t	Sig.
	B	Std. Error	Beta		
1	(Constant)	-13.672	5.293		
	pre-test score	.086	.005	.631	17.304
	chatterbooks	-1.031	.634	-.066	-1.626
	chatterbooksplus	-.085	.644	-.005	-.133
	AgeattestinMonths	.085	.035	.085	2.409
	school1	.255	1.358	.008	.188
	school2	2.218	1.053	.111	2.107
	school3	-1.260	1.348	-.041	-.934
	school4	-2.407	1.300	-.081	-1.852
	school5	.397	1.072	.019	.370
	school6	-1.642	1.314	-.055	-1.250
	school7	2.058	1.459	.059	1.410
	school8	-2.380	1.406	-.071	-1.693
	school9	-3.171	1.469	-.091	-2.159
	school10	-3.151	1.325	-.103	-2.377
	school11	-2.715	1.137	-.116	-2.389

a. Dependent Variable: post-test score

Neither female nor FSM was significant so these were excluded from the model.

Results of FSM model:

Coefficients ^a				
Model	Unstandardised Coefficients		Standardised Coefficients	
	B	Std. Error	Beta	t
				Sig.

		B	Std. Error	Beta		
1	(Constant)	-.750	3.272		-.229	.819
	pre-test score	.081	.010	.618	8.450	.000
	chatterbooks	-2.437	1.346	-.138	-1.811	.073
	chatterbooksplus	-.813	1.317	-.047	-.618	.538
	school1	-3.324	3.187	-.081	-1.043	.299
	school2	4.678	2.099	.256	2.229	.028
	school3	1.874	3.840	.036	.488	.627
	school4	-2.035	2.676	-.066	-.761	.449
	school5	4.397	3.554	.097	1.237	.219
	school6	.565	3.028	.015	.187	.852
	school7	5.531	4.573	.087	1.209	.229
	school8	-.171	2.791	-.005	-.061	.951
	school9	-2.739	2.504	-.101	-1.094	.277
	school10	-1.685	2.863	-.048	-.589	.557
	school11	-1.377	2.210	-.065	-.623	.535

a. Dependent Variable: post-test score

Appendix B: Contract for Schools

Please sign both copies, retaining one and returning the second copy to Prof. Clare Wood at Coventry University (Psychology Department), Priory Street, Coventry, CV1 5FB.

Agreement to participate in the Evaluation of 'Extra Curricular Reading Group Approaches to Improving Reading for Pleasure in Year 7 Pupils'

School Name: _____

Aims of the Evaluation

The aim of this project is to evaluate the impact of **Extra Curricular Reading Group Approaches to Improving Reading for Pleasure in Year 7 Pupils**. This project will evaluate the potential of two different 'reading group' approaches to improving reading and spelling outcomes for Year 7 children who have failed to reach Level 4 in English prior to secondary school. The approaches will be based on intensive versions of the Reading Agency's (a UK registered Charity) 'Chatterbooks' reading groups, which are designed to improve children's motivation to read for pleasure. One approach will follow the existing Chatterbooks format, and a second approach ('Chatterbooks Plus') will incorporate age-appropriate dialogic prompts to further develop the children's comprehension of the texts. Both approaches will be compared to a waiting list control group to see if they are able to improve performance in reading and spelling generally, and have boosted the children's motivation to read.

The results of the research will contribute to our understanding of what works in raising the pupil's 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 with reading difficulties.

The Project and Its Evaluation

A minimum of 450 Year 7 children who failed to achieve Level 4 in English at the end of Year 6 will be recruited in March 2013 and will complete: The New Group Reading Test (online) and the Single Word Spelling Test. These measures provide detailed information on the children's literacy abilities relative to UK norms. The children will also provide information about the children's reading behaviours outside of school and the children will complete an attitudes to reading measure. The children at each school will be randomly allocated to one of the three groups for a ten-week period. One group (minimum 150 children) will participate in a weekly Chatterbooks group. Chatterbooks is an extracurricular reading initiative designed by the Reading Agency. The focus of sessions is on creative and imaginative engagement around texts that are age-appropriate and pleasurable. Book choice is left up to the individual children each week. Another 150 children will be allocated to a 'Chatterbooks Plus' group, in which the children complete the Chatterbooks activities, but also engage in 15 minutes of dialogue around their chosen books with a focus on discussing relevant vocabulary and situational knowledge to enhance comprehension skills. Child attendance and engagement with the interventions will be monitored. The final 150 children will be allocated to a control group who will receive no treatment during the 10-week period. Once the results indicate which treatment appears to be most effective initially, the control group children will be offered the opportunity to participate in that approach early in Year 8. Both interventions will be administered as small-group sessions (approximately 15 children) led by a trained adult as an after-school 'club'. All books will be available to the children through local and school library services.

The children will be reassessed on the literacy and reading for pleasure metrics at the end of the Year 7, to evaluate the immediate impact of the approaches. Delayed post-tests will be administered at the beginning of Year 8, to determine whether the interventions can 'inoculate' children against the normal losses in literacy performance observed over the summer holidays.

The evaluation is being conducted by the National Foundation for Educational Research (NFER). Pupils who are selected and agree to take part are randomly allocated to either one of the two intervention groups or a waitlist control group. Random allocation is essential to the evaluation as it is the only way that we can say for sure what the effect of the intervention is on children's attainment. It is important that schools understand and consent to this

process. Members of the project team who will conduct the intervention will be randomly allocated to school sites and to one of the two intervention formats. They will be 'blind' to (unaware of) the format of the other intervention, and we ask all school colleagues not to disclose details of this to project team members. Colleagues from the NFER will also collect data on the implementation of the interventions, including observation of some sessions. Pupils' test responses and any other pupil data will be treated with the strictest confidence. With respect to the reading data, the responses will be collected online by GL Assessment and accessed by NFER. Named data will be matched with the National Pupil Database and shared with Coventry University and EEF. No individual school or pupil will be identified in any report arising from the research.

All pupils in the evaluation will be tested for improvement in reading and spelling outcomes as indicated in the project description above. In addition, the children will be asked to complete an 'attitudes to reading' questionnaire and provide some data on their reading activities outside of school.

RESPONSIBILITIES

THE TEAM FROM COVENTRY UNIVERSITY WILL:

- DELIVER 10 INTERVENTION SESSIONS AND ASSOCIATED MATERIALS TO PUPILS IN THE INTERVENTION GROUPS
- BE THE FIRST POINT OF CONTACT FOR ANY QUESTIONS ABOUT THE EVALUATION
- ENSURE ALL STAFF CARRYING OUT ASSESSMENTS ARE TRAINED AND HAVE RECEIVED CRB CLEARANCE
- ENSURE ALL STAFF CONDUCTING THE INTERVENTIONS ARE TRAINED AND HAVE RECEIVED CRB CLEARANCE
- COLLECT THE SPELLING AND READING MOTIVATION / READING HABITS DATA FROM THE PROJECT
- PROVIDE ON-GOING SUPPORT TO THE SCHOOL
- SEND OUT REGULAR UPDATES ON THE PROGRESS OF THE PROJECT THROUGH A NEWSLETTER.

THE NATIONAL FOUNDATION FOR EDUCATIONAL RESEARCH WILL:

- CONDUCT THE RANDOM ALLOCATION
- COLLECT AND ANALYSE THE READING DATA AND PROCESS EVALUATION DATA FROM THE PROJECT
- ENSURE THAT READING ATTAINMENT DATA IS AVAILABLE TO TEACHERS TO DOWNLOAD
- ENSURE ALL STAFF CARRYING OUT OBSERVATIONS AND WORKING WITH PUPIL DATA ARE TRAINED AND HAVE RECEIVED CRB CLEARANCE
- DISSEMINATE RESEARCH FINDINGS.

The School will:

- Consent to random allocation and commit to the outcome (whether treatment or control)
- Allow time for each testing phase and liaise with the evaluation team and the project team to find appropriate dates and times for testing to take place
- Ensure the shared understanding and support of all school staff for the project and personnel involved
- Be a point of contact for parents / carers seeking more information on the project.

We commit to the Evaluation of Extra Curricular Reading Group Approaches to Improving Reading for Pleasure in Year 7 Pupils as detailed above

Signatures

ON BEHALF OF COVENTRY UNIVERSITY

PROJECT LEADER [PROF. CLARE WOOD]: _____

DATE: _____

ON BEHALF OF THE NATIONAL FOUNDATION FOR EDUCATIONAL RESEARCH:

LEAD EVALUATOR [BEN STYLES]: _____



DATE: 6/2/13

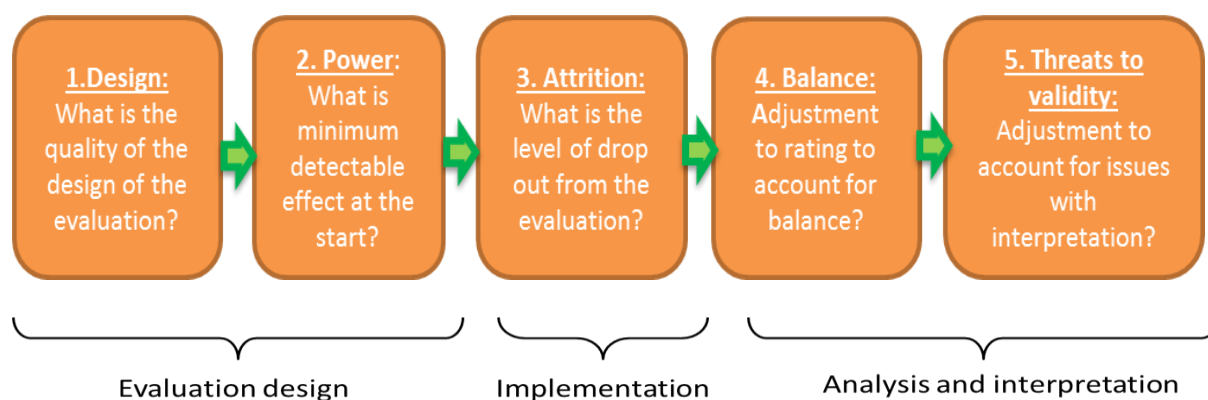
ON BEHALF OF THE SCHOOL:

HEADTEACHER [NAME]: _____

OTHER RELEVANT SCHOOL STAFF [NAMES]: _____

DATE: _____

Appendix C: Security rating summary - Chatterbooks



Rating	1. Design	2. Power (MDES)	3. Attrition	4. Balance	5. Threats to validity
🔒🔒🔒🔒🔒	Fair and clear experimental design (RCT)	< 0.2	< 10%	Well-balanced on observables	No threats to validity
🔒🔒🔒🔒🔒	Fair and clear experimental design (RCT, RDD)	< 0.3	< 20%		
🔒🔒🔒🔒🔒	Well-matched comparison (quasi-experiment)	< 0.4	< 30%		Some threats
🔒🔒🔒🔒🔒	Matched comparison (quasi-experiment)	< 0.5	< 40%		
🔒🔒🔒🔒🔒	Comparison group with poor or no matching	< 0.6	< 50%	↓	↓
🔒🔒🔒🔒🔒	No comparator	> 0.6	> 50%	Imbalanced on observables	Significant threats

The final security rating for this trial is 3 🔒. This means that findings are reasonably secure.

The trial was designed as a well-powered, individually randomised, efficacy trial and could have achieved a maximum of 4 🔒. It was run a short time period in only 12 schools, which is a concern for the generalisability of the findings.

The trial had attrition of 21% which is slightly above average. However, the groups were balanced on observables at the baseline after attrition. There was some threat to the validity of the study because the tests were not delivered blind by the evaluator, although test administrators did not have a role in intervention delivery. Also, because randomisation occurred within the school there could be some contamination, although this is unlikely because the intervention was not amenable to this kind of transfer.

Additional detail about the EEF's security rating system can be found at:

www.educationendowmentfoundation.org.uk/evaluation.

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v2.0.

To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/version/2 or email: psi@nationalarchives.gsi.gov.uk

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

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



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