Education Endowment Foundation

GraphoGame Rime

Evaluation report and executive summary

May 2018

Independent evaluators:

Jack Worth, Julie Nelson, Jennie Harland, Daniele Bernardinelli, Ben Styles

NFER

National Foundation for Educational Research

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.

This project was funded as part of the Education and Neuroscience scheme, which was jointly funded by the Wellcome Trust and Education Endowment Foundation and launched in January 2014. The aim of the scheme was to provide funding for collaborative projects between educators and neuroscientists to develop evidence-based interventions for use in the classroom, or to rigorously test existing tools and practices.

For more information about the EEF or this report please contact:

Danielle Mason
Head of Research and Publications
Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP
t: 020 7802 1679
e: danielle.mason@eefoundation.org.uk
w: www.educationendowmentfoundation.org.uk

About the evaluator

The project was independently evaluated by a team from the National Foundation for Educational Research: Dr Ben Styles, Jack Worth, Michael Neaves, Daniele Bernardinelli, Julie Nelson, Jennie Harland, Stephen McNamara and Lisa O'Donnell.

The lead evaluator was Dr Ben Styles.

Contact details:

Dr Ben Styles
Research Director
National Foundation for Educational Research
The Mere
Upton Park
Slough
Berkshire
SL1 2DQ

Email: b.styles@nfer.ac.uk Tel: 01753 637386

Contents

Executive summary 4

Introduction 6

Methods 10

Impact evaluation 18

Process evaluation 29

Conclusion 36

References 38

Appendix A: EEF cost rating 39

Appendix B: Security classification of trial findings 40

Appendix C: Memorandum of Understanding 41

Appendix D: Parent Letter and Information Sheet 44

Appendix E: Randomisation Syntax 47

Appendix F: Process evaluation achieved sample 48

Executive summary

The project

GraphoGame Rime is a computer game designed to teach pupils to read by developing their phonological awareness and phonic skills. Originally developed by a Finnish University, the program underpinning the game can analyse performance and constantly adjust the difficulty of the game content to match the learner's ability. The English version of GraphoGame Rime was developed to implement the 'rhyme analogy' research by educational neuroscientist Usha Goswami, based at the University of Cambridge. The intervention aimed to improve the reading ability of a group of pupils who were identified as having low literacy skills, as measured by the phonics screening check taken at the end of Year 1. The intervention was delivered by teachers and teaching assistants. Training, technical support and some delivery support was provided by researchers from the University of Cambridge.

This study was an efficacy trial of the GraphoGame Rime intervention, carried out during one spring term with 398 Year 2 pupils in 15 primary schools in Cambridgeshire. It was a pupil-randomised controlled trial designed to determine the impact of the intervention on the pupils' reading attainment and was conducted with two cohorts of pupils over two years. The primary outcome was the raw score on the New Group Reading Test, administered by NFER test administrators within a month of the intervention ending, and provided by GL Assessment. A process evaluation used case-study visits, telephone interviews and analysis of data on pupils' usage of the game to capture the perceptions and experiences of participating teaching staff and pupils. The intervention was implemented in schools during the 2015/16 and 2016/17 academic years.

GraphoGame Rime was funded as part of the Education and Neuroscience scheme, a collaboration between EEF and Wellcome Trust to provide funding for collaborative projects between educators and neuroscientists.

Key conclusions
1. The trial found no evidence that GraphoGame Rime improves pupils' reading or spelling test scores when compared to business-as-usual. This result has very high security.
2. The same is true when looking specifically at pupils who have ever been eligible for FSM. The security of this result is lower because the number of pupils is smaller.
3. Teachers reported that they felt sufficiently well trained and found the intervention easy to set up and implement. Teachers, senior leaders and pupils considered GraphoGame Rime highly engaging, motivational and enjoyable. Findings suggest that all schools implemented the programme with a relatively good level of fidelity.
4. Because the game was tested against business as usual, comparison group pupils received other literacy support, including small-group and one-to-one literacy activities, for similar amounts of time to that spent on GraphoGame Rime by pupils using it. This means the lack of observed impact shows that the intervention is no more or less effective than the support the comparison pupils received.

EEF security rating

The findings have very high security. The trial was a two-armed pupil-randomised controlled trial, which had a relatively large number of pupils (398 randomised; 362 in the final analysis). As an efficacy trial, it tested whether the intervention can work with delivery and support direct from the developer, and was tested in a relatively small number of schools in a small geographical area. The final analysis had sufficient statistical power to detect an effect size of 0.25, although the power was lower than expected at the design stage because the correlation between pre- and post-test scores was lower than

anticipated. The final analysis included primary outcome data for pupils from all 15 schools involved in the trial. Less than ten per cent of participating pupils had missing data because of absence on the day of testing or having left the school. This attrition is likely to be unbiased.

Additional findings

The evidence from our impact evaluation suggests the GraphoGame Rime intervention had no impact on pupils' reading attainment compared to a business-as-usual control. The analysis also suggests the intervention had no impact on pupils' spelling attainment and no impact on FSM-eligible pupils' reading or spelling attainment compared with the business-as-usual control.

Data on the number of hours pupils spent using the game shows that the level of implementation was generally good across schools. The developer's initial recommendation was that pupils should spend between 8.3 hours and 12.5 hours playing the game in total. In the trial, the average playing time was six hours in the first year and nine hours in the second year. 'On-treatment' analysis, which explored the association between the number of hours each pupil spent using the game and their test scores, did not show a relationship between more playing time and better outcomes.

The intervention was targeted at pupils who had scored below the Government's expected standard on the phonics screening check in Year 1. Government policy encourages this group of pupils to be supported by their school to improve their decoding skills, and the process evaluation found that the business-as-usual control pupils were experiencing a range of activities with similar aims to GraphoGame Rime, including additional literacy support in small-groups and one-to-one activities. Therefore, it could be the case that both the intervention and control group made accelerated progress in literacy compared to the rest of their class, but that GraphoGame was, on average, no more effective than the other interventions put in place to support pupils with identified decoding skill needs.

Cost

The estimated average cost per school per year for the intervention as delivered in this trial was around £770, or £48 per pupil per year when averaged over three years. The main financial costs of delivering the intervention were the costs associated with delivering training and on-going technical support. In addition, participating schools needed to allow half a day of time and supply cover per staff member, to allow the staff to attend the induction and training session (though these were also used for explaining the RCT evaluation requirements). Participating schools also required PCs, laptops or tablets and headphones to allow pupils to play the game without disturbing their classmates.

Table 1: Summary of impact on primary outcome

Group	Pupil numbers	Effect size (95% confidence interval)	P value	Estimated months' progress	EEF security rating	EEF cost rating
Treatment vs. control	362	-0.06 (-0.23 0.12)	0.48	-1	🔒🔒🔒🔒🔒	**£**£££
Treatment FSM (eligible in last 6 years) vs. control	109	-0.01 (-0.43 0.42)	0.98	0		**£**£££

Introduction

Intervention

GraphoGame Rime is a computer game developed to teach pupils to read by developing their phonological awareness. Originally developed by a Finnish University, the computer programs that are part of the GraphoGame group employ algorithms that analyse a child's performance and constantly adjust the difficulty of the content so that the challenge matches the learner's ability. The English version of GraphoGame Rime was developed on the basis of the rhyme analogy research carried out by educational neuroscientist Usha Goswami. The intervention has a number of features, which are outlined below.

Rationale: A computer game, requiring minimal supervision, which is played by pupils with the intention of improving their phonological awareness and reading skills. It has been developed by Usha Goswami and her team at Cambridge University, who supported the technical implementation of the software in participating schools (see pages 23-24).

Recipients: Year 2 pupils were identified for the trial using scores on the Year 1 phonics screening check – see Participants section (p.13) for more detail.

The resource: The game was initially available for PC and, in the second year of the evaluation, for tablet as well. It consists of 25 levels, with each level taking approximately 10-15 minutes to complete. Each pupil plays using a unique identifier (avatar) and headphones. The game supports pupils to match auditory patterns with groups of letters (rimes) that are displayed on the screen as multiple choice options. The game first focuses on rimes that are most common in English and offers increasingly challenging levels as pupils improve their skills.

Procedures and dosage: The principle is that pupils play the game for around 10-15 minutes each day, across all five days of the week for the whole of the spring term (i.e. approximately 10-12 weeks). The game is intended to be played during normal literacy class time, replacing normal literacy activities, rather than being additional to these. In a small number of cases, however, we found evidence that the game was played outside of normal literacy time (see page 28 for details). GraphoGame pupils should play the game in a discreet classroom, or area of the classroom.

Implementers: The playing of GraphoGame should be supervised by teachers or teaching assistants (TAs) to ensure that children are able to log on, and to make sure that they remain on task throughout the session and do not collaborate with their peers. There is no requirement for active tuition by teachers or TAs.

Training: Teachers were invited to a single half-day training session at the beginning of Year 1 of the intervention. The session was located at the University of Cambridge and led by Usha Goswami and others from her development team. The session focused on: introducing GraphoGame and its key principles – including a demonstration; explaining how schools should implement the game correctly as part of the RCT (in terms of appropriate supervision; correct dosage; and ensuring that data from each playing session was recorded correctly); and details of the technical support that would be provided by the Cambridge team and the IT systems that schools needed to have in place. Cambridge University staff provided technical assistance to schools throughout the course of the trial, but especially in Year 1, in response to a number of initial technical problems encountered by schools (see page 23). The technical problems arose because the software development was still being finalised when the trial began. These problems were largely overcome in Year 2. Training in Year 2 was delivered as needed, for individual schools.

Monitoring and Quality Assurance: The team from Cambridge remotely monitored game usage data throughout the trial and notified schools that were below the recommended usage time. The team maintained contact with the schools throughout the trial to check fidelity to the trial requirements.

Background evidence

The issue of how best to teach reading in the early years of education has divided researchers and education professionals for decades (see, for example Wyse and Styles, 2007). Since Sir Jim Rose's Review of the Teaching of Early Reading (Rose, 2006), schools in England have been required to support pupils to read through the use of systematic synthetic phonics. This is a method that focuses on teaching children to recognise individual phonemes (letter sounds) which are matched to appropriate graphemes (letters), and then blended (so that, for example, "c"-"a"-"t" = cat). Irregular words (those which cannot be pronounced phonically) are introduced slowly and systematically once regular code is learned.

The neuroscientist Usha Goswami argues that this approach is not always appropriate, particularly in the context of a language such as English, which is highly irregular:

> *English is an exceptionally inconsistent alphabetic language... It is relatively easy to learn about phonemes if one letter consistently maps onto one and the same phoneme, or if one phoneme consistently maps to one and the same letter. It is relatively difficult to learn about phonemes if a letter can be pronounced in multiple ways (e.g. the letter "A" in English maps onto a different phoneme in the highly familiar words "cat", "was", "saw", "made", and "car")* (Goswami, 2005).

Goswami's research on reading using "rhyme analogies" (light-fight) shows that children's learning is helped by a focus on 'rime' units in rhyming families of words (Goswami, 1986). In this approach, children are taught to recognise the initial consonant or consonant blend in a syllable (the onset) and the vowel/consonant combination that follows that (the rime) (so, for example "c", "at" = cat). If a word has two syllables, it has two rimes. Words that have the same written 'rime' pattern (for example, "c-at", "b-at", "m-at") also 'rhyme' (that is, they 'sound' the same) – this link between sounds and written words (through rhyme and rime) is a key focus of GraphoGame Rime. Goswami provides the example of the rime spelling "ight". She notes that there are 90 words in the English language with this rime spelling, and that "pronunciation is consistent across all of them." She comments: "This makes it likely that English children need to develop phonological recoding strategies at more than one "grain size" [phoneme level] in order to become competent readers" (Goswami, 2005).

Not all words that 'rhyme' will necessarily have the same written 'rime' pattern, however (for example, bl-ue, gr-ew, sh-oe, thr-ough). Children progress onto these more complicated 'rimes' once the basic principles are learned.

GraphoGame Rime (GraphoGame) is a computer game that has been developed to teach pupils to read through the use of the rime method. Originally developed in Finland, the game employs algorithms that analyse a child's performance and constantly adjusts the difficulty of the content so that the challenge matches the learner's ability. An English version of GraphoGame has been developed by Usha Goswami's research team.

A review of neuroscience and education for the Education Endowment Foundation (EEF) found moderate evidence that "computer-based training focused on phonological skills has helped those experiencing difficulty to develop their reading skills" (Howard-Jones, 2014). Two small, non-randomised, studies showed promising results. One Finnish-based study showed that just three hours of use improved reading-related skills, while another, which compared the direct translation of GraphoGame into English with a version that additionally introduced the largest rhyme families with the

most consistent orthographic rime spellings, found that both games led to gains in reading, spelling, and phonological skills (Howard-Jones, 2014). Both studies were small in scale and had limitations, however, and, as a result, the EEF has commissioned a further trial to test these results. GraphoGame Rime has been the subject of two published small non-randomised studies carried out in Cambridge, which suggest promising effect sizes on specific measures such as spelling (0.9) and phonemic awareness (over 1). The first study (N=31) compared playing GraphoGame Rime to playing an alternative English language version of GraphoGame, GraphoGame Phoneme, the latter developed in Finland (Kyle et al., 2013). Small groups played either GraphoGame Rime or GraphoGame Phoneme daily for 12 weeks, or formed an untreated control. The GraphoGame Rime group showed medium effect sizes for reading (0.66, 0.53) and large effect sizes for spelling and phonic decoding (0.91, 1.43). The GraphoGame Phoneme group showed small effect sizes for reading (0.22, 0.43) and spelling (0.45) and a medium effect size for phonic decoding (0.60). The second study (N=19) compared playing GraphoGame Rime to an oral intervention developed by the Cambridge researchers based on poetry and music. The poetry/music intervention was designed to enhance phonological skills and thereby reading (Bhide et al., 2013). Gains from GraphoGame Rime (pre-test to post-test) showed large effect sizes (reading: 2.03; phonic decoding: 1.28; spelling: 1.4). Note, however, there was no unseen control group in this study. A reading intervention is usually considered effective if effect sizes are greater than 0.13 – 0.23 (Torgesen et al. 2001).

Neural imaging indicates that there are also brain signatures (“neural markers”) that reflect the success of reading interventions. Small-scale studies show that playing GraphoGame improves these neural markers in children receiving the intervention. For example, the N170 is a negative deflection in brain electrical activity that typically occurs 170ms after the onset of a letter string, if the letter string is a real word. The N170 is thus considered a neural marker for word-specific neural processing, across languages. The German version of GraphoGame has been shown to foster the development of an N170-like response in Swiss-German kindergarten children (Brem et al., 2010). Control children who played a closely-matched maths software game intervention did not show this neural response, suggesting that it indexes the development of a specialized letter-processing system for reading words. fMRI data (a neural imaging measure based on changes in blood flow) taken at the same time showed that the Swiss-German children receiving GraphoGame, but not the control children, also showed greater neural activation in the Visual Word Form area (VWFA) following the intervention. The VWFA is an area of visual cortex that becomes specialised for word recognition. Brem et al.’s data thus suggest that GraphoGame training was working specifically on the neural areas specialised for reading. Meanwhile, the Finnish version of GraphoGame has been shown to improve a neural marker for phonetic processing, the electrical mis-match negativity response (MMN). Playing GraphoGame for 3 hours increased the size of the MMN to different Finnish phonemes in Finnish-speaking children, but not in control children playing the maths software game. The degree of MMN increase also correlated with the degree of children’s improvement in letter knowledge (Lovio et al., 2012). Neural imaging studies of the English language GraphoGame Rime have not been carried out.

Evaluation objectives

The primary research question was whether playing GraphoGame Rime as part of literacy lessons improved pupils’ reading ability, compared to ‘business as usual’ participation in literacy lessons. Secondary research questions of the impact evaluation include the impact of the intervention on pupils’ spelling ability, and the impact on the sub-sample of disadvantaged pupils (who have been eligible for free school meals at any time during their schooling).

The process evaluation aimed to explore the effectiveness of the implementation of the intervention and provide insight into how it was used and received by schools. The process evaluation collected information on:

- preparation of staff to deliver the intervention

- implementation of GraphoGame Rime, including challenges, adaptations, and fidelity to the delivery plan
- resources involved in the intervention, including training and staffing costs (e.g. use of teaching assistants, cover time for training)
- perceived outcomes and impacts
- deliverers' views on the suitability, sustainability, and potential for roll out of the intervention
- pupils' views about the intervention.

EEF published the evaluation protocol, which set out the above objectives, when the project began in October 2015[1]. An amended protocol was published in March 2016[2], reflecting changes to the evaluation as a result of the first pre-test. The subsequent deviations from the protocol are highlighted in the report alongside an explanation of why the deviation occurred.

Ethical review

The evaluation and consent procedure was approved by University of Cambridge's ethics board as well as by NFER's Code of Practice committee. University of Cambridge obtained headteacher agreement to participate in the trial when the school was recruited. A memorandum of understanding setting out the central commitments and data-sharing details that were required was shared with, and signed by, the headteacher (see Appendix C). Parents were made aware of their right to opt-out of their child's data being linked to the NPD or used in any way in the research via a letter to the parents of eligible pupils, sent out by the schools (see Appendix D).

Project team

The project was led by Professor Usha Goswami at the University of Cambridge and managed by Mary Anne Wolpert, with support from Dr Anji Wilson, Hannah Noble and Henna Ahmed.

The evaluation was overseen by Dr Ben Styles at NFER. The impact evaluation was led by Jack Worth with assistance from Michael Neaves and Daniele Bernardinelli. The process evaluation was led by Julie Nelson, assisted by Jennie Harland, Stephen McNamara and Lisa O'Donnell.

Trial registration

The trial was registered at the ISRCTN registry in September 2015[3].

[1] **https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/graphogame-rime/**
[2] **https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/graphogame-rime/**

[3] **http://www.isrctn.com/ISRCTN10467450**

Methods

Trial design

The evaluation design was a pupil-randomised controlled trial with two arms: a group of pupils that spend 10-15 minutes each day playing GraphoGame Rime computer games during literacy lessons and a 'business as usual' control group of pupils from the same classes who will have business-as-usual tuition in literacy lessons. A pupil-randomised design was preferred to a school- or class-randomised design in order to maximise statistical power, given the restraints on the number of schools that the delivery team could work with. The risk of contamination was small as the intervention could be used by intervention pupils with minimal risk of being used by control pupils in their class. The intervention being tested was use of the computer game, so teachers' training about the intervention was, on its own, unlikely to have an influence on control pupils' learning.

The randomisation was stratified by class so that the number of pupils that were in the intervention and control groups, respectively, were known by teachers long in advance, while the identity of the pupils that are allocated to each group was not revealed until after the pre-testing had been completed to prevent any bias. The trial was conducted using two cohorts of pupils over two years. The number of schools needed to be relatively small to support delivery, but the trial was two years instead of one to ensure the number of pupils was sufficient for the required statistical power. The protocol specified an interim analysis at the end of the first year and associated stopping rule, in the event of a positive or negative effect being found.

Participant selection

Eleven primary/infant schools in Cambridgeshire with Year 2 pupils were recruited by the University of Cambridge team in the first year of the trial (i.e. 2015). The trial was conducted over two years with the aim of including all the same schools in both years. The school decided how to allocate their staff, whereas pupils were randomised (see p.9).

Cohorts of Year 2 pupils from these schools participated in the trial. One school dropped out after the completion of the post-testing in the first year and a further four schools were recruited to participate in the second year. The first-cohort pupils in the school that dropped out had been pre- and post-tested and pupils that would have been in the second cohort in that school were not pre-tested or randomised. Therefore, the fact the school dropped out of participating in the trial was not attrition from a measurement or evaluation perspective.

Year 2 pupils that scored below 32 marks out of 40 (the government's expected standard) on the phonics screening check as Year 1 pupils were eligible for participation in the evaluation.[4] This threshold was chosen to target the programme at struggling readers and to ensure that a consistent selection threshold was applied across all the schools involved. However, the fixed threshold meant that there was some variety between schools in terms of pupil numbers: some had as many as 20 per class and others just one or two eligible pupils.

Outcomes measures

The University of Cambridge team administered the New Group Reading Test (NGRT) Level 1B test as a pre-test measure in both years of the trial. The originally planned primary outcome measure was the raw score on the NGRT Level 2. However, during the administration of the GraphoGame Rime pre-test in the 2015 autumn term (during the first year of the trial) the Cambridge team expressed concern that pupils were finding the reading pre-test (NGRT Level 1B) very difficult. The marked test score data confirmed this: the average raw score (9.7) was only just above that expected from guessing (8.6) and

[4] The phonics screening check is a statutory assessment of decoding ability taken by all Year 1 pupils in England.

the scores had a low variance. A low variance of scores (i.e. the test not discriminating between pupils of different abilities) would be detrimental to the trial analysis, and pupils struggling with a test that is beyond their capabilities could be distressing for them.

All parties, including the EEF, therefore agreed that the protocol be amended to make the raw score on the NGRT Level 1B test the primary outcome measure of the trial, instead of the NGRT Level 2 test. Using the same test at post-test as was used at pre-test might have the disadvantage that pupils remember some of the content, which could affect average scores. However, this would not necessarily introduce any bias, as all pupils (intervention and control) sat the same pre-test. We tested for the interaction between pre-test and intervention (see Outcomes and analysis section (p.17)). While this was based on a pre-test that was misaligned in terms of ability, it did measure something and the result was that there was no interaction.

The developer also hypothesised that the intervention would improve pupils' spelling ability, so we measured the impact on spelling by administering the GL Assessment Single-Word Spelling Test (SWST) as a post-test. We used raw scores from this 30-item test in the analysis.

The paper versions of the NGRT and SWST were used, provided by GL Assessment. The tests were administered in small groups by independent NFER test administrators, who did not know the group allocation of the pupils being tested and were requested not to try to find out.

The protocol stated that another secondary outcome would be "subject to availability through the National Pupil Database, the spelling component of the end of Key Stage 1 grammar, punctuation and spelling (GPS) test". However, this was not used as schools were not required to administer the GPS test after the test's confidentiality was breached (DfE, 2016), no data for Key Stage 1 tests was collected (only teacher assessments) and the teacher assessment judgements did not distinguish between spelling and the other components.

Sample size

The evaluation aimed to recruit ten primary or infant schools, and have an average of 20 pupils in each school in each of the two years of the study. The power calculations were therefore based on the assumption of 400 pupils participating in the trial, allocated half to the intervention group and half to the control group.

We assumed a correlation between the pre-test reading measure and the outcome measure would be around 0.8[5], meaning the design would be able to detect an effect size of 0.17 with 80% power. Attrition of ten per cent would still yield a minimum detectable effect size at 80% power of 0.18. The intra-cluster correlation was assumed to be zero for the power calculations because randomisation occurs within schools.[6]

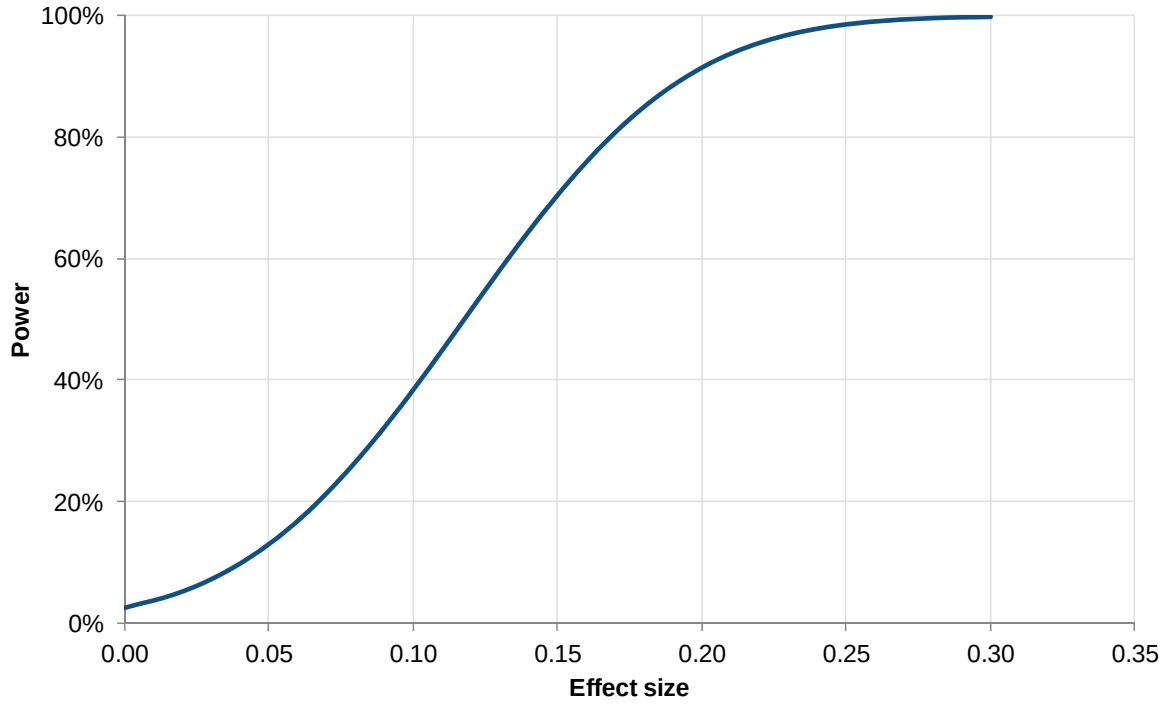
The sample of schools was a small convenience sample and it is therefore not appropriate to generalise the findings to a wider population of schools. Analysis used a single-level regression model with dummy variables for class, to take into account the restricted randomisation. Figure 1 plots the statistical power the design has to detect different effect sizes, given the assumptions made. It shows that the design had a good chance (more than 80 per cent, a conventional threshold) of detecting effect sizes larger

[5] This was based on the correlation in Worth *et al.,* (2015), using reading tests at the beginning and end of Year 2. This study differed in that it used a different reading test (New Group Reading Test rather than Progress in English) and a subset of the class was tested rather than the whole class.

[6] We analysed this trial with the assumption of 'conditional inference' (Hedges and Vevea, 1998) i.e. inference is restricted to the schools and pupils within the study. This meant the most appropriate analysis model treated school as a fixed effect. This in turn meant that sample size calculations that are based on simple randomisation (with no ICC) are conservative. Modelling school as a random effect would have required a much larger sample (and a non-zero ICC).

than 0.17, but less statistical power to detect smaller effects. The chance of detecting an effect size below 0.12 using this design is less than half.

Figure 1: Power calculations



Randomisation

Each eligible pupil who had completed a pre-test was allocated to either the intervention or control group. An independent, but unblinded NFER statistician used SPSS software to perform a stratified randomisation of pupils by class, with the aim of balancing the randomisation at class level (i.e. having equal numbers in intervention and control groups within each class) to aid delivery. We first put schools in a random order; within school we put classes in a random order; and within classes we put pupils in a random order. The number of classes per school varied from one to three. The entire pupil list was then allocated to control and intervention in turn: first pupil to control, second to intervention, third to control, and so on. The SPSS syntax used to randomise pupils in the first year is shown in Appendix E, and similar syntax was used in the second year.

Analysis

Primary analysis

We have conducted the analysis in line with EEF's analysis guidance (EEF, 2015). The primary analysis used a linear regression model of reading test score at post-test including all data from both cohorts. The model was estimated with all pupils that completed a pre-test and a post-test. The primary outcome model had raw score at post-test as the dependent variable and the following covariates were included in every model:

- an indicator of whether the pupil is in the GraphoGame intervention group. The excluded group was the control group, so the coefficient of the intervention group indicator measured the difference in (conditional) outcomes between the intervention group and the control group
- the pupil's raw score on the pre-test. The coefficient is incidental to the research, but explained a large proportion of outcome variance, increasing the power of the analysis.

- an indicator of which class and intervention cohort the pupil was in, to account for the stratified randomisation.
- As pupils are nested within classes and classes are nested within schools, an indicator for school was not necessary for this single-level model. Therefore, we modelled class as fixed effect for the primary model. There is considerable debate in clinical trials about how to report the main intervention effect when there is a significant site-by-treatment interaction (Feaster *et al*, 2011). For the purposes of this study, the main effect was estimated from a model without interaction terms i.e. the weighted mean of treatment effects across sites.
- The evaluation protocol specified gender and age in months as additional covariates, but these were not included in the final analysis to align with EEF's analysis policy. These variables would explain a small amount of variance in addition to pre-test score, so would be of limited value.

The coefficient on the intervention group indicator measures the impact of the intervention, measured in terms of raw test score. As is standard for EEF evaluations, the standardised effect size was calculated as the coefficient on the intervention group indicator, divided by the total sample standard deviation. We applied Hedges' adjustment for small sample bias of estimated variance.

In addition, a further model explored the interaction between class and intervention in order to determine whether GraphoGame was differentially effective between different classes or schools. This analysis was expected to be somewhat underpowered, but it was anticipated that it could yield interesting results in terms of whether certain classes or schools had larger effects than others.

A planned interim analysis took place at the end of the first year on just the primary outcome. The interim results were discussed with EEF to determine whether or not the trial should continue, given that substantial evidence of a positive effect of the intervention might have led EEF to advise moving to a larger effectiveness trial, while evidence of negative effects might have led EEF to advise stopping the trial. However, it was known that the analysis of data from the first year was underpowered for plausible positive or negative effects, so it was most likely that the trial would be completed as planned.

Undertaking interim analysis that influences whether or not the final analysis will be conducted affects the way inference ought to be made. Armitage et al. (1969) were the first to numerically compute the extent to which the type I error probability is increased over its nominal level if a standard hypothesis test is conducted at each of a series of interim analyses. They studied the problem of testing a normal mean with known variance and set the significance level (or type I error probability) for the trial to be 5%. If one interim analysis and one final analysis are performed this error rises to 8%. This was the situation here. Although the final result was not predicated on a significance level, the decision at the interim analysis was binary and we determined that we should therefore adopt an appropriate threshold to aid our decision. The threshold $p < 0.0294$ is the Pocock boundary for two planned analyses (Pocock, 1977). We therefore also used $p < 0.0294$ (and a 97% confidence interval) for the results of the final analysis. Any reference to the trial demonstrating a 'significant' result is still at the conventional level of $p < 0.05$.

Sensitivity analysis

We carried out analysis of the primary outcome using a multilevel model (as opposed to a single-level model with class fixed effects) with class and class-by-intervention interaction as random effects. The analysis was conducted to test whether the class-by-intervention interaction was significant at the 5 per cent level. If not, the effect size itself (without the interaction term) should be the same as that from the fixed effects model. If significant, this would tell us something about the extent to which the effectiveness of the intervention varies depending on which class it is delivered in. Were we trying to generalise to a wider population of schools, this would be important to take into account. Since we are not seeking to generalise results beyond the schools in this efficacy trial, the effect size and confidence interval from the fixed effects model (without interaction) serve as the definitive result of the trial.

Sub-group analysis

The primary analysis models were estimated for the sub-sample of pupils that have been FSM since starting school (the indicator of Pupil Premium status). The effect size was measured in the same way as above. We re-estimated the standard deviation for the FSM-only sample to use to standardise the estimated effect size. We also estimated an FSM-by-intervention interaction model to analyse the differential impact of the intervention on FSM pupils.

We also estimated the primary analysis model with the addition of an interaction between pre-test score and the intervention group indicator, to explore differential impacts of the intervention on pupils with higher or lower levels of ability, as measured at pre-test.

Secondary analysis

We estimated the primary analysis models with Single Word Spelling Test raw scores as an outcome variable to assess the impact of the intervention on pupils' spelling ability.

Exploratory analysis

We conducted exploratory analyses that focussed on the implementation of the intervention to attempt to understand *how* the intervention worked, rather than whether it worked. Differences in the way the intervention was implemented across schools were at the discretion of school staff and researchers on the Cambridge team, so this analysis cannot be definitively interpreted as causal. However, the analysis may give clues as to what drives any effect that may be identified by the impact analysis. It investigated whether different approaches to implementation led to different amounts of progress made by pupils.

We conducted an on-treatment analysis of the number of hours spent using GraphoGame. Pupil-level game usage data were recoded into levels of programme fidelity using a correspondence that was pre-agreed with the developer and published in the protocol (shown in Table 8). We explored the differential impact of the intervention by fidelity level by including indicator variables for each fidelity level as well as the intervention group indicator. We analysed the data without recoding, to give an indication of the amount of impact an additional hour of using GraphoGame made to reading.

Whilst this analysis appears attractive, it is very vulnerable to bias as those individuals who used the program the most are likely to have other characteristics that are associated with improved test performance. To mitigate this, Complier Average Causal Effect analysis (Sussman and Hayward, 2010) was carried out using a continuous variable to describe dosage.

Implementation and process evaluation

Eleven schools participated in Year 1 of the trial (2015-16) and a further four joined the trial in Year 2 (2016-17) – 15 schools in total. The process evaluation aimed to explore the experiences of these schools across the two years, in particular to ascertain whether the intervention was implemented with fidelity in terms of:

- **dosage**: Whether pupils were playing GraphoGame for the recommended 8.3 to 12.5 hours over the course of each year of the trial[7]
- **timing of delivery**: Whether pupils were playing GraphoGame during usual literacy lesson time
- **control-group activity**: Whether control pupils were engaged in matched-time literacy activities while intervention children were playing GraphoGame
- **contamination**: Whether there was any distraction or contamination between intervention and control group pupils.

[7] Between 10 and 15 minutes, five days a week for ten weeks.

Fidelity was agreed in consultation with the delivery team and the EEF prior to analysis and included in the Statistical Analysis Plan. Further details about interviews conducted through the process evaluation are described below, and are outlined in greater detail in Appendix F. Appendix F also provides details of the code assigned to each school for reporting purposes

The process evaluation was commissioned to evaluate the experiences of *ten* schools that were initially signed up to implement GraphoGame across each year of the trial. In the event, eleven schools were recruited in Year 1. In Year 1, the process evaluation team (which was independent of the impact evaluation team) collected information from ten of these 11 schools (we were unable to secure an interview or an observation with school J, although this school remained in the trial). In Year 2, a further four schools joined the trial, but school D dropped out, so the number of participating schools was now 14. We collected information from ten of these 14 schools (six of the Year 1 schools plus all four of the Year 2 schools) in the second year of the trial. We made the decision to replace four of the original schools with new schools to explore whether the technical challenges experienced by the Year 1 schools were common only in the first year of the trial, or whether they were also issues for the new (Year 2) schools.

In Year 1, we conducted five half-day case-study visits (with schools A-E) and five lighter-touch 30-minute telephone interviews (with schools F-I and K). We purposefully selected schools that had made varying degrees of progress in each group. The five case study- and five telephone-interview schools each contained a mixture of schools that had not yet implemented the game, those that were experiencing difficulties with implementation, and those that were implementing the game with relative ease. This information was provided by the developer.

- Each of the half-day case-study visits was conducted in the spring term and included: an observation of pupils using GraphoGame; face-to-face interviews with teachers (or other staff involved in supervising the implementation of the intervention) and a senior leader; and paired or small-group discussions with Year 2 pupils who were involved in the intervention. In cases where the schools would only accommodate one interview, we interviewed just one person, always a teacher (Appendix F provides details). We conducted a brief 15-minute follow-up telephone catch-up with the teacher from each of these schools in the summer term.

- Each of the five light-touch telephone interviews was conducted early in the summer term with the teacher responsible for the implementation of GraphoGame in his or her respective school. In just one case (School G) the person interviewed was a senior leader.

In Year 2, we visited three of the original case-study schools again (schools A, B and C) for a second half-day case-study visit and we conducted a second light-touch telephone interview with three of the original telephone interview schools (schools G, H and I). We did not visit or conduct an interview with schools E, F or K again, because these schools had proven difficult to engage in the first Year of the trial (and school D had dropped out of the trial by this point). Although difficult to engage in the evaluation, we have no reason to believe that these schools were any less engaged in the intervention itself than other schools. Instead, we replaced schools D, E, F and K with the four schools that were new to the trial in Year 2. Schools L and O took part in a detailed case-study visit, while schools M and N took part in a light-touch telephone interview. All case-study visits took place in the Spring term, and telephone interviews were conducted in the Summer Term.

In total, we collected data from 14 schools over both years of the trial. We conducted 27 teacher interviews, seven senior leader interviews, and nine observations of GraphoGame in use. We also carried out short paired or small-group discussions with 29 Year 2 pupils to explore their experiences of the game (Appendix F provides further details). Interview and observation data was coded and analysed thematically, in line with the topics covered on pages 25-31 of this report.

In addition to the case-study and interview data, we received quantitative data from the programme developers about the amount of time that each pupil spent playing GraphoGame and the highest level of difficulty the pupil progressed to. This usage data was collected from school computers and was linked to pupils' test scores via their individual login details for exploratory on-treatment analysis.

Costs

We gathered information about the cost of the intervention from the University of Cambridge and information about any additional costs incurred by schools from teachers during the case-study visits and telephone interviews. This information consisted of the overall costs for delivering training and supporting implementation, additional time or financial costs borne by schools, and any pre-requisite costs. We established, from the nature of each activity, which costs would be one-offs and which would be on-going, if a school continued with the intervention over a number of years.

We estimated a per-school cost based on the number of schools that were in the evaluation and divided the cost per school by the average number of eligible pupils in a school. We profiled the cost over three years by considering the one-off and on-going costs separately, as per the EEF guidance.[8]

Timeline

Table 2 shows the timetable of delivery and evaluation activities that took place over the duration of the trial.

[8] **https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EEF_guidance_to_evaluators_on_cost_evaluation.pdf**

Table 2: Timeline

Date	Activity
April – June 2015	Recruited schools
July – October 2015	Collected pupil data from schools (first year)
September 2015	Induction day for teachers and staff
October – December 2015	Pre-testing of pupils in first year of trial by University of Cambridge
December 2015	Randomisation of pupils in first year of trial by NFER, schools informed
January - April 2016	Intervention implemented Process evaluation case-study visits
April – May 2016	Post-testing by NFER test administrators Process evaluation follow-up telephone interviews
July – October 2016	Additional school recruitment for second year and pupil data collection from schools (second year)
August 2016	Interim analysis (impact evaluation only) conducted by NFER
September – December 2016	Pre-testing of pupils in second year of trial by University of Cambridge
December 2016	Randomisation of pupils in second year of trial by NFER, schools informed
January - April 2017	Intervention implemented Process evaluation case-study visits
April – May 2017	Post-testing by NFER test administrators Process evaluation follow-up telephone interviews
August – November 2017	Final analysis conducted by NFER. Evaluation report written.

Impact evaluation

Participants

Figure 2 shows the participant flow diagram for the trial, combining the numbers of Year 2 pupils from both cohorts that participated. The pupil eligibility criteria meant that the number of eligible pupils was, on average, a small proportion of the pupils within each class. On average, 7.5 pupils per class were eligible for the trial and 3.8 pupils per class were randomised to receive the intervention. However, the number of eligible pupils per class varied between 1 and 20, with a median of 7. Typical class sizes are around 30 pupils, meaning that typically around a quarter of pupils were eligible, whereas in some classes as many as two-thirds were eligible.

The diagram shows that pupil-level attrition from the study was limited to less than ten per cent overall. The main reasons for pupil drop out were pupils being absent or unavailable on the day of testing (26 pupils) and a small number having left the school (eight pupils). Attrition was fairly balanced across intervention and control pupils (seven per cent in the intervention group and eleven per cent in the control group), and there is no reason to suspect it was biased. Three pupils were randomised in error: their phonics check score was coded as zero in the data collected from schools (suggesting they were eligible) but they had been absent or disapplied from the phonics check (and were in fact ineligible). These pupils were excluded from the analysis.

As mentioned above, one school dropped out of the trial after the first year had been completed (the randomised pupils had done the intervention and been post-tested) and before the second year had begun (the second cohort of pupils was not pre-tested or randomised). This is not included in the diagram as it does not constitute pupil attrition. In addition, four new schools were recruited to the trial for the second year only.

Figure 2: Participant flow diagram

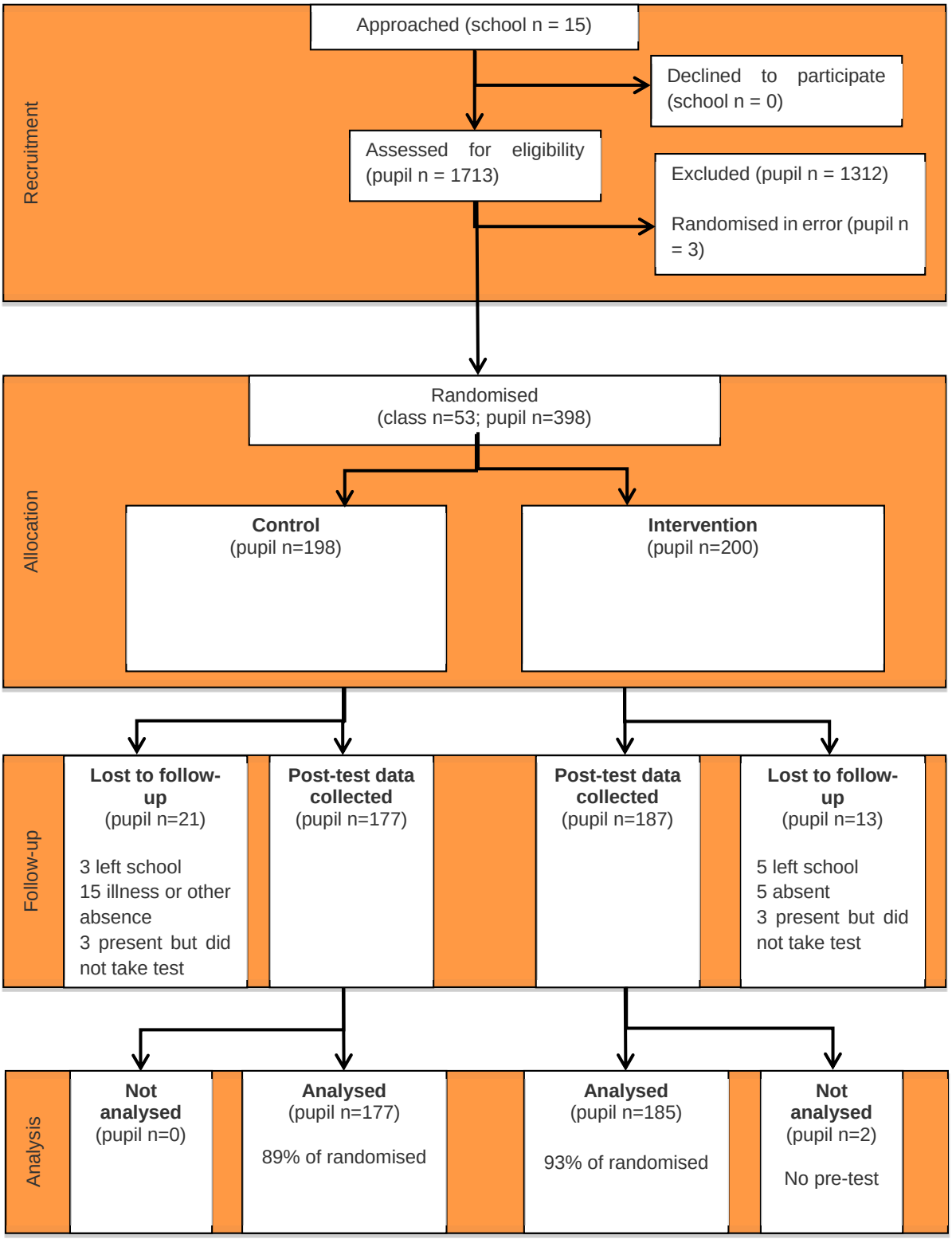


Table 3 shows the sample size calculations made at different stages of the trial. The trial was designed to achieve a minimum detectable effect size (MDES) of 0.17. School and pupil recruitment was good, so the statistical power at randomisation reflected the power at the design stage, subject to the underlying assumptions being correct. However, one of our underlying assumptions was not correct: the correlation between pre- and post-test was lower than expected, due to the difficulty of the pre-test. Despite pupil attrition being low, the lower correlation cost the design in terms of statistical power, implying a MDES of 0.25 for the final analysis.

Table 3: Minimum detectable effect size at different stages

Stage	N pupils (n= intervention; n=control)	Correlation between pre-test (+other covariates) & post-test	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)
Protocol	400 (200; 200)	0.80	zero	No stratification assumed	80%	0.05	0.17
Randomisation	401 (200; 201)	0.80	zero	No stratification assumed	80%	0.05	0.17
Analysis (i.e. available pre- and post-test)	362 (185; 177)	0.57	zero	No stratification assumed	80%	0.05	0.25

Pupil characteristics

The trial involved 15 primary schools across the two years, with 11 in the first year and 14 in the second year. One school from the first year did not participate again in the second year, when four new schools were recruited.

The characteristics of the participating schools are reported in Table 4 below, with a comparison to schools in England, where relevant. Schools were not selected to create a random sample, and there is no intention to generalise the results of the trial beyond the participating schools. The majority of schools involved in the trial had a proportion of FSM pupils higher than their respective local authority average and all but two of the schools have relatively low Key Stage 2 attainment, with the proportion of pupils achieving the expected standards below the national average.

Table 4: School characteristics

Variable		Schools in trial		Schools in England
School-level (categorical)		**n**	**Percentage**	**Percentage**
School type	Community school	9	60%	40%
	Foundation school	3	20%	4%
	Voluntary aided	0	0%	17%
	Voluntary controlled	2	13%	11%
	Academy	1	7%	27%
Local authority	Cambridgeshire	13		
	Peterborough	1		
	Hertfordshire	1		
Ofsted	Outstanding	0	0%	19%
	Good	9	60%	71%
	Requires Improvement	6	40%	9%
	Inadequate	0	0%	1%
School-level (continuous)		**n (missing)**	**Mean**	
Average number of Y2 pupils on roll		15(0)	57	41.2
Average percentage of FSM pupils in the school		15(0)	14.3%	14.8%

Table 5 below shows some key characteristics of the pupils involved in the trial, among pupils randomised and in the analysis sample. Given the small extent of attrition between pre- and post-tests, the pattern of differences are very similar in the randomised and final analysis samples of pupils. The randomisation process ensured that there was no systematic difference between the intervention and the control group, although some small imbalances might be expected due to random variation. There were indeed some small differences in prior attainment between the intervention and the control group, as measured by the phonics screening check and the reading pre-test. The confidence intervals show that these differences were not statistically significant and, in the case of pre-test score, were taken account of in the analysis by inclusion as a covariate.

Table 5: Baseline comparisons

Randomised sample

Variable	Intervention group		Control group		Difference
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**	**(percentage points)**
Female	78/198 (2)	39.4%	73/197(1)	37.1%	2.3
Eligible for FSM in the past six years (ever6)	69/197 (3)	35.0%	56/195 (3)	28.7%	6.3
Pupil-level (continuous)	**n (missing)**	**Mean**	**n (missing)**	**Mean**	**Effect Size (95% CI)**
Phonics Mark	196 (4)	18.1 (8.7)	192 (6)	19.4 (8.9)	-0.15 (-0.34 0.04)
Pre-test score (NGRT 1B)	198 (2)	8.6 (4.3)	198 (0)	8.9 (5.3)	-0.09 (-0.27 0.09)

Analysis sample

Variable	Intervention group		Control group		Difference
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**	**(percentage points)**
Female	73/184 (1)	39.7%	63/176 (1)	35.8%	3.9
Eligible for FSM in the past six years (ever6)	63/183 (2)	34.4%	47/175 (2)	26.9%	7.5
Pupil-level (continuous)	**n (missing)**	**Mean (SD)**	**n (missing)**	**Mean (SD)**	**Effect Size (95% CI)**
Phonics Mark	182 (3)	18.5 (8.6)	173 (4)	19.8 (8.7)	-0.14 (-0.34 0.06)
Pre-test score (NGRT 1B)	185 (0)	8.7 (4.3)	177 (0)	9.1 (5.3)	-0.08 (-0.27 0.11)

Outcomes and analysis

Table 6 below presents a summary of the results from the analysis of both primary and secondary outcomes, reporting the estimated effect size of the difference between the intervention group and control group, and the relative confidence intervals.

Table 6: Summary of impact analysis

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)[1]	p-value
Reading	185 (15)	13.3 (12.2 - 14.4)	177 (21)	13.9 (12.8 - 15.1)	362 (185; 177)	-0.06 (-0.23 0.12)	0.482
Spelling	183 (17)	16.7 (15.8 - 17.7)	177 (21)	16.8 (15.7 - 17.8)	360 (183; 177)	0.01 (-0.18 0.19)	0.953
Reading (FSM ever 6 only)	63 (6)	12.6 (10.5 - 14.7)	47 (9)	11.5 (10 - 12.9)	110 (63; 47)	-0.01 (-0.43 0.42)	0.980
Spelling (FSM ever 6 only)	62 (7)	16.1 (14.5 - 17.7)	47 (9)	15.1 (13.3 - 16.9)	109 (62; 47)	0.05 (-0.43 0.52)	0.844
Interim analysis (Reading - first year only)	98(10)	15.4 (13.8 – 17.1)	100 (8)	16.1 (14.5 – 17.8)	198 (98; 100)	-0.04 (-0.29 0.20)	0.714

Note: [1] for the primary (reading) and interim analyses, the confidence interval is a 97% confidence interval. See Analysis section for detailed explanation of why this is the case.

The primary analysis consisted of a linear regression model with post-intervention reading test scores as the dependent variable, controlling for pre-intervention test scores, group allocation, and class fixed effects. The estimated effect size as a proportion of the standard deviation was **-0.06 (97% C.I. -0.23 0.12)**. The primary analysis suggests that the intervention had no statistically significant impact on pupils' reading ability.

As part of the pre-specified secondary analysis, the primary regression model was run with post-intervention spelling test scores as the dependent variable. As with the primary outcome, the estimates suggest the intervention had no impact on spelling, with an effect size of **0.01 (95% C.I. -0.20 0.21)**.

The same analysis for reading and spelling outcomes was performed on the subgroup of pupils eligible for free school meals (EVERFSM6). The intervention group post-test mean was higher than in the control group, but the difference was reduced after controlling for pre-test scores and class fixed effects with our regression model. These estimates suggest the intervention had no impact on the reading or spelling ability of disadvantaged pupils (effect sizes and confidence intervals are reported in Table 6).

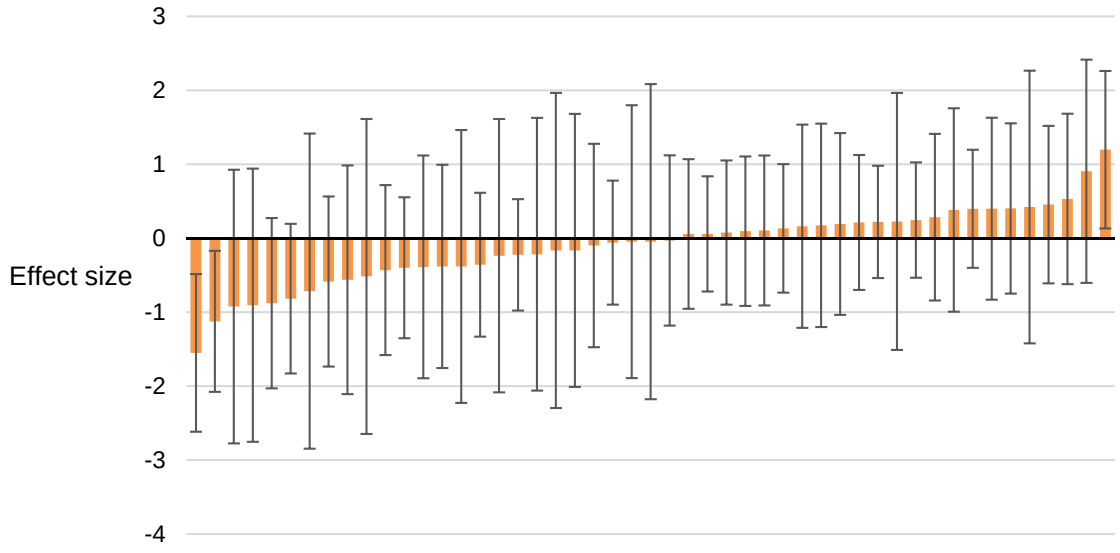
Additionally, the primary regression model was run including the interaction between pre-test scores and the group allocation (see Table 7 below). The interaction term would capture a possible differential effect of the intervention for pupils with different pre-test scores. The estimated interaction parameter was not significantly different from zero, suggesting the impact of the intervention did not systematically differ among pupils with different levels of prior attainment.

Table 7: Interaction between group allocation and pre-intervention test scores

Variable	Effect Size (95% CI)	P-value
Pre-test score	0.10 (0.08 0.13)	0.000
Intervention group	-0.01 (-0.35 0.34)	0.971
Interaction	**-0.01 (-0.04 0.03)**	**0.744**

Finally, as part of the primary analysis, a further model with interaction terms between class indicators and the intervention indicator was conducted, to assess the differential effect of the intervention between classes. Figure 3 below shows the variability of the impact across classes, and as only one confidence interval did not overlap with zero there was no evidence of a differential effect of class. However, this analysis is underpowered, due to the substantial number of units (53 classes across the two years) and corresponding interaction terms, relative to the number of pupils.

Figure 3: Estimated class-specific effect sizes for the primary outcome with 95% confidence intervals



Note: Interaction results for 50 classes are shown rather than 53. This is because (1) the largest class was excluded as the base category, and (2) two classes were excluded because they had two pupils, precluding estimation of an interaction term.

As part of the pre-specified exploratory analysis, we analysed the relationship between the number of hours pupils spent using the GraphoGame software, and reading test scores. The number of hours spent by each pupil was grouped into one of four levels of programme fidelity, which were pre-defined by the developer and included in the evaluation protocol. Table 8 below shows how the usage data was categorised into fidelity levels. Data on the number of hours was available for 183 of the 185 intervention pupils included in the primary and secondary analysis.

Table 8: Classification of program fidelity for the intervention group

Level of programme fidelity	Total number of hours spent on GraphoGame Rime	Number of pupils
Low	0-3	1
Medium	3-6	77
High	6-10	73
Very high	More than 10	32

We found a negative correlation between the amount of time pupils spent using GraphoGame and reading post-test scores: this suggests that pupils who spent longer playing the game were pupils who made less progress in reading. This difference also remains after taking account of pupils' pre-test scores. This does not necessarily imply that spending more time on the game caused less progress as measured by the test scores. The amount of time spent using the game was a choice made by pupils and/or teaching staff, so was not randomly assigned and could reflect other underlying differences between the pupils/ teachers.

Indeed, we found that pupils who spent more time on playing the game had lower pre-test scores, suggesting that lower-performing pupils tended to spend more time on the game. As indicated in Table 9, this seems to be confirmed when looking at the highest level of difficulty each pupil reached within the game: the correlation between the difficulty levels reached and test scores shows a weak but positive relationship. This suggests that pupils who struggled more with reading before the trial started spent more time playing the game than their classmates, but did not progress as far into the game's difficulty levels.

Table 9: Correlation between game participation, levels of difficulty and test scores

Programme data	Correlation with pre-intervention test score (p-value)	Correlation with post-intervention test score (p-value)
Hours spent playing the game	-0.168 (0.022)	-0.298 (0.00)
Highest difficulty level reached	0.051 (0.494)	0.116 (0.115)

Further exploratory analysis on the relationship between programme fidelity and attainment was carried out by extending the regression model to include levels of usage and highest difficulty levels reached instead of the intervention group indicator variable. Due to the small numbers of observations, the Low and Medium usage levels were grouped together. We also grouped the 181 sub-levels of game difficulty into 8 levels for analysis (based on the assessment levels within the game). Table 10 below shows the results of our regression analysis.

The estimates in Table 10 show a negative relationship between usage of the computer game and reading scores, after controlling for pre-test scores and class-fixed effects. Pupils in the higher usage groups tended to have lower reading scores. However, none of the parameters are statistically significant. On the other hand, the analysis based on game-difficulty levels shows that pupils who reached higher difficulty levels tended to have higher reading scores. The two lowest groups had a negative and statistically significant effect size.

Table 10: Association between levels of attainment and levels of programme fidelity

Fidelity level	Group N	Effect Size (95% confidence interval)
Low/Med	77	0.08 (-0.15 0.31)
High	73	-0.11 (-0.34 0.12)
Very High	32	-0.31 (-0.66 0.04)
Assessment Level	**Group N**	**Effect Size (95% confidence interval)**
Below assessment level 3	17	-0.52 (-0.92 -0.11)
Completed assessment level 3	39	-0.29 (-0.57 0)
Completed assessment level 4	39	-0.16 (-0.44 0.11)
Completed assessment level 5	36	0.28 (-0.01 0.57)
Completed assessment level 6	23	0.04 (-0.32 0.4)
Completed assessment level 7	31	0.14 (-0.19 0.47)

The analysis shows that pupils who spent more time playing the game than their classmates, and those who did not progress as far into the game's difficulty levels, made slightly less progress with reading during the trial, although the relationships are not statistically significant. Intervention pupils with lower pre-test scores were the ones who tended to spend more time playing the game and tended to progress less far up the difficulty levels, suggesting that the differences are likely to be driven by underlying factors affecting the reading progress made by that group of pupils (for example, greater developmental needs) more than they are by greater exposure to the game.

Cost

The GraphoGame Rime intervention was implemented as part of an EEF project rather than as an intervention that schools paid for. The cost of the intervention, presented below, has been estimated on the basis of costs that were part of the project that was delivered. However, the costs do not necessarily represent what the cost would be to a school if it were to deliver this intervention independently. For example, the cost of a school receiving individual training sessions may be higher than if a number of schools attend the training, as happened in the trial. It is not necessary to attend training to use the game, as teachers can download a manual from the Cambridge University website, but the cost of the induction training has been included in the cost estimate because it is likely to have impacted on implementation during the trial.

This cost evaluation estimates the cost to schools in terms of financial costs, pre-requisite costs and compensation for staff time if there was no EEF trial in place. We have assumed that the computer game software would be made available free of charge. If this is not the case then the cost for accessing the software would be additional to the costs identified below.

Financial costs

The main financial costs of delivering the intervention, which were borne by EEF as part of this project, were the costs associated with delivering training and supporting implementation in schools. The cost per school of training is a one-off cost and our estimate of the cost per school is based on the cost and number of schools (11) in the first year of the trial. Some top-up training was delivered in the second year for the four additional schools, but the most realistic setting for delivering the training was the first year when it was spread among the eleven schools. Table 11 presents the one-off costs associated with delivering training, which average £261 per school.

Table 11: University of Cambridge's financial costs for delivering training

Cost item	Total cost	Cost per school per year
Preparing training materials, photocopying and delivery team time spent delivering training	£1,500	£136
Administration time arranging training with schools and room booking for training	£774	£70
Travel	£600	£55
Total	**£2,873**	**£261**

Note: Due to rounding, the sum of the individual amounts may not exactly sum to the total.

The cost of University of Cambridge research assistants supporting schools with implementing the intervention can be considered an ongoing cost for schools (see 'Implementation' section below). The cost of supporting implementation was greater in the first year of the trial than in the second year, because most technical problems with the software had been resolved by the second year and schools could use remote support rather than hands-on support. The final version of the game can be used without support from the University of Cambridge research assistants. Our estimate of the cost per school is based on the average cost and number of schools across the first and second years of the trial (11 in the first year and 14 in the second year). Table 12 presents the ongoing costs associated with implementation support, which average £683 per school per year.

Table 12: University of Cambridge's financial costs for implementation support

Cost item	Year 1 cost	Year 2 cost	Average cost per school
Research assistant staff time supporting delivery in schools	£9,750	£3,438	£528
Research assistant travel costs for school visits	£1,274	£1,142	£97
Administration support for delivery in schools: arranging visits, photocopying	£525	£945	£59
Total	**£11,549**	**£5,525**	**£683**

Note: Due to rounding, the sum of the individual amounts may not exactly sum to the total.

During interviews as part of the process evaluation, we asked senior leaders and teachers whether their school had incurred any additional financial costs to implement GraphoGame. Most schools reported that there had been no *additional* financial costs to deliver GraphoGame during the trial. The only exception was that one school had purchased two headphone splitters so that the TAs supporting pupils one-to-one to use the game could also hear its auditory signals (approximate cost £10).

We combined the one-off financial costs (training) in the first year with the ongoing costs (support) in the first and subsequent years to model the total cost per school over three years. This reflects the cost of implementing the intervention over a sustained period. Table 13 shows our estimate of the cost per school over time, the cumulative cost per school over three years, and the average cost per school per year. The average cost per school per year over three years was £770.

We derived the cost per pupil by dividing the average cost per school per year by the average number of eligible pupils per school per year. Based on an average of 16 pupils per school per year, the average

cost per pupil per year over three years was £48. This cost is classified as very low according to EEF's cost ratings (see Appendix A).

Table 13: Average financial cost per school per year

Cost item	Cost per school	Cumulative total cost per school	Average cost per school per year	Average cost per pupil per year
First year	£944	£944		
Second year	£683	£1,627		
Third year	**£683**	**£2,310**	**£770**	**£48**

Staff time

As part of the trial, at least one member of staff from each school took part in training with the programme developers, which took approximately half a day for each staff member involved. It is not necessary to attend training to use the game, as teachers can download a manual from the Cambridge University website.

The time required to implement GraphoGame in schools was predominantly considered to be part of staff's normal teaching and learning responsibilities and, although time spent on this had supplanted other activities, it usually did not require any *additional* time investment. School staff spent time on activities such as:

- familiarising themselves with the game (approximately 2 hours)
- setting the game up on the schools' computer facilities and dealing with any technical issues (approximately 1-2 hours plus ongoing time as required)
- setting up computers/ tablets and logging pupils on and off the game (approximately 10 minutes per session)
- supervising pupils playing the game.

Prerequisite costs

Pupils required headphones to be able to play the game in class without distracting other children. These were provided by the University of Cambridge as part of the trial. The cost of providing headphones to all pupils in the first year was £1,622, an average of £15 per pupil. Computers or tablets were also considered essential equipment to play GraphoGame, but schools in the evaluation already had these.

Process evaluation

Introduction

The process evaluation investigated how the intervention was delivered, to provide contextual information to inform the impact findings, and to check fidelity to the trial requirements. The process evaluation involved consultations with 14 of the 15 schools taking part over the course of the two year trial. In the following sections we describe and discuss:

- the implementation of GraphoGame (including the conditions for success and challenges encountered)
- fidelity to the trial requirements, control group activity, and contamination issues
- perceived outcomes
- formative findings (including the attractiveness of the intervention to stakeholders and participants' suggested improvements).

Implementation

Training, preparation and support

Overall, schools felt well prepared and sufficiently trained to deliver the trial and use GraphoGame. Schools either attended the induction training held at the University of Cambridge or, if they were unable to attend, a member of the Cambridge team delivered training at their school site. The consensus was that training was effective. When asked about preparedness to implement the trial, one teacher commented: '*We knew exactly what the format was going to be, it was very clear, there wasn't any fuzziness*'. Many teachers felt confident enough to train up their Teaching Assistants (TAs) to run the intervention themselves. Schools involved in both years of the trial did not find it necessary to send individuals for further training or re-training following the first year.

In spite of a general sense of preparedness, some schools did encounter teething problems in preparing for the trial.

- **Technical challenges**. Schools involved in the first year of the trial were much more likely than those involved in the second year to encounter technical problems. In Year 1 there were numerous issues, particularly in the first few weeks of implementation, related to: poor compatibility of the game with schools' security and systems; glitches in the game causing it to freeze and crash; updated versions of the game being launched mid-way through the trial; and the game not always saving properly (for example, only saving to computer hard-drives rather than network servers, or not recording how long a pupil had spent on the game). Issues with GraphoGame not automatically recording how long pupils played were reported in four of the ten schools consulted in Year 1. To mitigate this challenge, University of Cambridge staff asked schools to keep a manual log of the time pupils spent playing the game and visited schools to access game usage data stored locally on computers. Such technical issues were quickly resolved within the first few weeks of implementation and were reflected in the calculations of how long pupils played the game (i.e. dosage). Schools that joined in the second year reported far fewer implementation problems. The technical issues encountered in Year 1 had, by then, largely been resolved.
- Schools that participated in both years of the trial felt more prepared in the second year than the first. This was due to a better understanding of what the trial entailed, and the benefit of lead-in time – having the entire autumn term to ensure that everything was adequately prepared. There were no reports of schools using GraphoGame with pupils prior to the trial

period. The teething problems encountered in the first year had largely been resolved by the second year.

Across all schools, the University of Cambridge was involved in the initial setting up of GraphoGame onto hardware devices and provided ongoing support to schools throughout the trial, as required. Schools widely praised the technical support provided.

- **Five schools involved in the first year of the trial** reported receiving regular, ongoing *hands-on* support from the Cambridge team to rectify technical teething problems and support the supervision of GraphoGame.
- **All four schools that became involved in the second year of the trial** required only *remote* support via email from the Cambridge team following initial set up.
- **Schools that were involved in both years of the trial** generally required *very little* support from the Cambridge team in their second year. By this point communication links between the University and individuals responsible for IT in schools had been formed. Additionally, schools had the whole of the autumn term to prepare and ensure that hardware and software were working effectively prior to implementation following Christmas in Year 2. The school which switched to tablets in the second year required some initial technical support from the Cambridge team, but this was relatively minor.

These points reflect the fact that most technical problems with the software had been rectified by the second year of the trial. They were not persistent problems requiring ongoing input from the Cambridge team, which might impact on longer-term scalability.

Organisation and Timing of Sessions

Schools involved in both years of the trial reported fewer organisational and timing issues in the second year compared to the first. This indicates that the issues raised in Year 1 were not pervasive.

The following aspects were noted in terms of delivery across year 1 and 2:

- **Timing and duration:** In most cases GraphoGame was played during literacy lessons either first thing in the morning or in the afternoon following lunch. In nearly all schools, children played for 10-20 minutes per session, approximately three to five times per week, with shorter sessions usually running more frequently. However, two schools ran daily 20-minute sessions. Another school noted that University of Cambridge staff notified them if they were below the recommended usage time and, in response, attempted to increase playing time. For schools which participated in both years of the trial, dosage and session lengths saw little change.
- **Timescale:** Most schools ran the trial from just after the Christmas holidays to the beginning of the Easter holidays. However in schools where the intervention began late, or which faced significant disruption mid-trial, GraphoGame was played for some additional weeks. One school, which started late but did not have the option of extending the intervention trial period past Easter, ran longer sessions of 35-40 minutes, supervised by Cambridge staff, to make up playing time.
- **Groups:** GraphoGame children usually played as one whole intervention group; with each child playing the game on an individual computer or tablet at the same time. However, in some circumstances this was not possible, for example if the intervention group was too large for available resources or if intervention children needed a high level of support. In these cases intervention children were split into smaller groups. Some schools with small group delivery encountered logistical challenges, for example, struggling to accommodate all intervention sessions over the time that had been allocated for a TA to supervise the intervention. In these instances, sessions sometimes had to be cut short to ensure that all children were able to play the game.

Supervision Arrangements

In most schools, supervision of pupils playing GraphoGame was provided by a TA, though in some cases it was a teacher, and in two schools a member of staff from Cambridge University. The majority of schools only required one teacher or TA to supervise intervention sessions, but where there were large intervention groups, or intervention groups that had children with high levels of need, these tended either to be split into multiple sessions with one TA, or were covered by multiple members of staff (e.g. both a teacher and a TA). Schools that ran multiple sessions with multiple members of staff encountered fewer supervision and organisational issues, such as cutting sessions short, than those which ran multiple sessions with one supervisor.

Supervision was mainly provided to help children with the log-on process and to ensure that they stayed on task. Some supervision arrangements provided additional support to children with greater needs such as those with motor control or hearing difficulties.

Equipment, resources and space

The majority of schools reported that they had sufficient equipment, resources and space to run the intervention, although schools with large GraphoGame groups experienced some logistical issues because IT suites were not always available. In one specific example (the same school in which Cambridge staff supervised the intervention in school due to TA sickness), the University of Cambridge provided laptops for the school during Year 1 of the trial, although this was not needed in Year 2.

Overall, schools involved in both years of the trial reported fewer issues regarding equipment, resources and space in the second year. Increased knowledge and preparation time allowed for laptops to be block booked and appropriate spaces confirmed for intervention use.

Children mostly played GraphoGame on PCs or laptops, although one school ran the game on tablets in the second year of the trial. Several teachers noted they would have preferred to run GraphoGame on tablets given the familiarity of many children with this technology. Interviewees reported that tablets were also more accessible for children with motor difficulties.

In terms of location, children usually played in an IT suite, but the intervention was also carried out in open-plan rooms, spare rooms and, occasionally, in Year 2 classrooms. All schools received headphones to play GraphoGame from the University of Cambridge.

Fidelity

Participants' understanding of the trial requirements

All schools reported a sufficient level of understanding of trial requirements to enable them to run the trial. However, there were some instances where interviewees would have appreciated further clarification. Several interviewees who were responsible for running the trial in school had not received all the information, about how to use the game and run the trial, in sufficient time, because a different member of staff had been involved initially and this information had not been effectively cascaded within school. Several teachers also raised other issues, including that they were not entirely clear on the rationale for the selection and allocation of pupils and that they were disappointed that children assigned to the control condition would not be able to play the game and wanted clarification on when this would be possible.

Fidelity and adaptation

Overall, there was a high level of fidelity between the intended and actual implementation of GraphoGame during the trial. This was chiefly because the computer game, once set up, was relatively easy to implement as a self-contained and standardised intervention, requiring minimal input, differentiation, or adaptation by teachers. Pupils were able to work through the programme content relatively independently and at their own pace.

However, three key implementation adaptations came to light through the evaluation:

1. **Dosage:** Analysis of the amount of time pupils spent playing GraphoGame indicates that in the first year of the trial, pupils in all 11 schools were *underexposed* to the intervention. The recommended playing time for the trial period was between 8.3 hours and 12.5 hours whereas, in practice pupils played the game for an average of six hours over the course of the trial. This aspect of the implementation improved substantially in the second year, with the average time spent playing GraphoGame going up to nine hours and the percentage of pupils in the "High" and "Very High" groups rising from 29% in the first year to 90% in the second.

   Teachers often attributed the underexposure in the first year to a delay of several weeks at the beginning of the trial due to technical challenges with the game, or other issues such as staff and pupil absence and special events in the school. Pupils' underexposure to the intervention may have moderated any impacts of the game. In addition, teachers explained that, for logistical reasons, pupils often played GraphoGame for slightly longer periods than the 10-15 minutes intended but for fewer sessions per week, rather than daily. This finding suggests that schools may find it challenging to use GraphoGame on a daily basis and may prefer to use it less frequently for slightly longer periods to warrant the timetabling, supervision, space, and equipment required.

2. **Mode of delivery:** Although most schools used GraphoGame during normal literacy lessons (such as guided reading, differentiated phonics-based activities or spelling activities), around one third played GraphoGame during non-literacy activities such as assembly, registration, other subjects or topic-based learning. This was sometimes for logistical reasons, such as a lack of staff availability to supervise pupils, or was due to a desire to avoid removing intervention pupils from core-literacy activities with their teacher. This suggests that a minority of pupils in the treatment group played GraphoGame *in addition to,* rather than as *part of*, their normal literacy lessons. Furthermore, because the pupils involved in this trial were identified as struggling with literacy, they also sometimes received other forms of literacy support (such as one-to-one reading or handwriting practice). That said, control pupils received a comparable amount of support (which, like the intervention pupils, sometimes included receiving their comparable activity during non-literacy time), so this should not undermine the impact evaluation findings.

3. **Level of difficulty:** GraphoGame is designed in such a way that pupils work through increasingly challenging content at their own pace. As each level of the game is successfully completed, new and more challenging levels become available. This feature means that we should expect some variation in the level of difficulty of the game experienced by pupils. However, two points are noteworthy:

   - Some pupils reportedly repeated levels. One teacher explained that: '*Some children were replaying levels because they felt that was easier than moving on*'. Over the course of the trial this did not seem to be a widespread problem. Pupils were generally motivated to progress through the game, and replaying was prevented by staff supervision where necessary, but it was an issue in four schools and was identified as one of the weaknesses of the game.

   - Three teachers thought that because all pupils entered the game at the same level of difficulty (regardless of their pre-existing abilities), some pupils had not been sufficiently challenged by the early levels of the game and therefore that it had offered little scope

for progress. The impact analysis found a negative interaction between pre-test score and the impact of the intervention (see Table 7), which implies that pupils with lower pre-test scores tended to show more progress as a result of the intervention. However, this result is very weak, and not statistically significant, suggesting that lack of differentiated challenge was not a significant general issue.

Outcomes

Teachers were overwhelmingly positive about the game. However, the vast majority were hesitant to specifically assign any improvements in phonics attainment or progress to playing it. Two notable exceptions were:

- one school, in which interviewees thought that GraphoGame had an effect on results in the phonics screening in both years of the trial
- another school which, in the second year of involvement in the trial, thought that GraphoGame had a profound effect on two children. One child was identified as a much higher attainer than previously thought. The other child saw a dramatic improvement in score when staff ran a mock phonics screening.

Teachers from five of the schools believed that playing GraphoGame increased children's confidence. The majority thought that children very much enjoyed playing it and that it was a good experience for them.

In most cases, the children who played GraphoGame also gave very positive feedback. The majority understood how to play it, reflecting the intuitive nature of GraphoGame noted by several teachers. Many of the intervention children thought that playing GraphoGame had improved their skills in writing, reading and spelling. The extent to which these self-perceived improvements were in one, a combination, or all, of these skills varied across the children. One child thought that GraphoGame would help her to '*read big books*'. A few children referenced the progressive nature of the game. As one said '*at the start it's a bit easier but then it gets a bit harder*'. Some liked this feature, possibly because they perceived themselves as improving. One child made reference to the fact that the technique presented in GraphoGame helped him when not using the game stating '*if there is a tricky word… we can chop it up into pieces and answer it*.' In endorsing the game, one child simply said: '*I love it, it's a game and I love games*'.

Formative findings

Attractiveness of the intervention to stakeholders

As indicated in the perceived outcomes section, teachers, senior leaders and pupils were overwhelmingly positive about GraphoGame. They considered it highly engaging, motivational and enjoyable. Respondents highlighted the following features of GraphoGame as being particularly effective (in order of the frequency mentioned):

- easy-to-follow instructions so pupils could play the game relatively independently with only minimal support from an adult
- progressive difficulty of activities meaning that pupils could progress through the games at their own pace and appropriate to their individual ability level
- varied and high quality graphics and activities
- rewards (such as virtual stickers and money to spend in a virtual shop) that pupils could access once they had completed a set amount of playing time
- instant feedback for pupils on whether they had correctly identified a sound or not

- emphasis on rhyming sounds providing an alternative approach to synthetic phonics programmes.

All but one of the teachers and senior leaders that we interviewed said that they would recommend GraphoGame to other schools and colleagues. These interviewees also wanted to continue using GraphoGame with the pupils in their school either as a targeted intervention, or for use with younger children to help strengthen their phonics ability earlier. Interviewees suggested that headphones, computers (or tablets) and moderate supervision were necessary if the game was to be used more widely in schools. One teacher explained why she would recommend GraphoGame:

> *'I would definitely recommend the game, if it came up for purchase I would buy it tomorrow. It's what's been lacking - a stretching programme for English and it does motivate them, it's not reliant on adult support, it's easy to use as an intervention tool and you can monitor progress easily through it*'.

Improvements to the intervention

Although interviewees were generally very positive about GraphoGame, they were not entirely without criticism. Four teachers questioned its capacity to develop pupils' phonetic understanding and reading ability overall; suggesting that it focused on a very specific aspect of phonics (i.e. a pupils' ability to identify rime units). While most teachers felt that GraphoGame provided a complementary and alternative emphasis in teaching pupils to read, one teacher was concerned that it was contradictory to synthetic phonics programmes (for example, by focusing on the identification of rimes rather than phonemes).

Interviewees identified some ways in which they thought GraphoGame could be improved to make it even more effective. Their suggestions are categorised below (in the order of frequency mentioned):

- tablet/touch screen version - this would provide greater flexibility in terms of where the game is played and eliminate any difficulties that pupils have with controlling a computer mouse
- reading strategies and context – the game could be extended to include alternatives to the rhyme analogy approach (e.g. synthetic phonics decoding), more context (to help pupils understand word meaning), and comprehension practice (e.g. embedding the correct word into a sentence) in order to support the development of broader literacy skills.
- individual pupil login – to reduce the amount of staff input required to set up the game
- initial screening of pupils – this would enable pupils to access the game at different levels rather than via a single entry point. Single point entry had the result that some pupils were not sufficiently challenged by the initial levels of GraphoGame
- locking of completed levels – so that pupils cannot repeat levels already played and must progress to more challenging content
- pupil performance analytics – this would enable teachers to explore how pupils are progressing with the game and to ascertain if there are any particular aspects they are struggling with, to inform further support. Although such analytics were actually a feature of the game, and teachers had been shown where to find them, this issue was raised by two interviewees.
- better style and presentation – the game could be made more familiar to pupils by adjusting aspects such as the font (for example, the formation of the letter 'a' was not compatible with the style that pupils were taught to use when forming this letter), and the pronunciation of words (the narrator's accent was unfamiliar to some children so they struggled to distinguish some sounds, for example, 'ing' and 'ink').
- clear explanation – to ensure that each sound encountered in an activity has been properly introduced and explained

- data synchronisation – regular auto-saving so that if the game crashes, pupils do not have to repeat completed activities.

Control group activity

Control group literacy support. Intervention and control pupils received different *types* of literacy support during the trial. While intervention pupils played GraphoGame, control pupils received alternative literacy support. Activities for control pupils mainly involved differentiated literacy activities with the class teacher as part of normal phonics or guided reading sessions, or small group/one-to-one support from a teacher or TA focusing on specific phonics content and skills (in four schools this involved use of phonics-based computer games).

Matched-time activity. Overall, pupils in the control group tended to be exposed to similar *amounts* of literacy support as pupils in the intervention group, and in most cases the support was delivered simultaneously. Both GraphoGame and control group pupils received a similar amount of further literacy support (i.e. beyond the specific and timed treatment/control activity), including tuition on synthetic phonics, spelling and writing. In a few cases, control-group activity was not simultaneous with the playing of GraphoGame. For instance, pupils took part in small-group or one-to-one activities at different times of the day due to staffing availability.

Control activity 'business as usual': Control group activities primarily comprised the schools' normal practices for pupils requiring additional literacy support to help them pass the phonics screening check. However, two schools had purchased new resources (a computer spelling game and a phonics board game) to make the activities more appealing and to assuage any feelings of resentment among control pupils.

Distraction. Where the intervention and control group activities took place in the same location there was quite often evidence of moderate distraction and of the control pupils being envious of pupils playing GraphoGame, particularly in the initial stages of the trial. As one teacher explained:

> '*They* [the pupils] *see it and they want a go. That was one of the hiccups at the beginning. It was hard for them to think 'well why are they on it and I'm not' and then there was a bit of 'I'm not going to do my work because I want to be on the computer'. So we've had to say 'these children are doing it now and there might be an opportunity for you to have a go later. That's put them at ease and they're used to it now*'.

In spite of some distraction issues, teachers reported that it was highly unlikely that there was any **contamination** between the intervention and control groups as it was not possible for non-intervention pupils to access GraphoGame. To do this they needed a teacher login and pupils were generally supervised throughout the activities. So, although there was some evidence of control pupils showing a degree of resentment that they were not able to play the game, and becoming distracted from their usual literacy activities as a result, we found no evidence that this resulted in contamination. Neither did we find evidence that the intervention drew staff away from control pupils, although a small number of interviewees indicated that it prevented staff from supporting one-to-one or small-group work with pupils who needed it.

Conclusion

Key conclusions
1. The trial found no evidence that GraphoGame Rime improves pupils' reading or spelling test scores when compared to business-as-usual. This result has very high security.
2. The same is true when looking specifically at pupils who have ever been eligible for FSM. The security of this result is lower because the number of pupils is smaller.
3. Teachers reported that they felt sufficiently well trained and found the intervention easy to set up and implement. Teachers, senior leaders and pupils considered GraphoGame Rime highly engaging, motivational and enjoyable. Findings suggest that all schools implemented the programme with a relatively good level of fidelity.
4. Because the game was tested against business as usual, comparison group pupils received other literacy support, including small-group and one-to-one literacy activities, for similar amounts of time to that spent on GraphoGame Rime by pupils using it. This means the lack of observed impact shows that the intervention is no more or less effective than the support the comparison pupils received.

Interpretation

The evidence from our impact evaluation suggests the GraphoGame Rime intervention had no impact on pupils' reading attainment over a business-as-usual control. This result is reasonably secure: the number of schools and pupils involved in the trial was as designed and the rate of attrition was low, although the low correlation between pre- and post-tests meant the statistical power of the final analysis was lower than anticipated at design stage. The pupil-randomised design means there could be threats to internal validity, although the process evaluation found that it was very unlikely that contamination had occurred.

Despite the security of the finding being lower than anticipated at the design stage, the amount of statistical precision the primary analysis has is enough to rule out effect sizes as large as those found in previous studies looking at GraphoGame's effects on spelling (0.9) and phonemic awareness (1.0).

Data on the number of hours pupils spent using the game shows that the level of implementation was generally good across schools. The level of implementation in the first year of the evaluation was below the developer's recommendations that pupils should spend between 8.3 hours and 12.5 hours playing the game, with pupils playing for an average of six hours. Incomplete implementation may therefore contribute to explaining why the evaluation found the intervention had no impact on pupil outcomes, although the level of implementation improved in the second year to an average of nine hours. On-treatment analysis, which explored the association between the number of hours each pupil spent using the game and their test scores, did not show a relationship between greater usage and improved outcomes on the NGRT.

In fact, pupils who spent relatively more time playing the game made slightly less reading progress than other pupils. These pupils tended to have lower pre-test scores and also tended to progress to lower levels of game difficulty, suggesting that underlying pupil ability was driving the relationships seen, rather than greater interaction with the game.

The intervention was targeted at pupils who had scored below the Government's expected standard on the phonics screening check in Year 1. Government policy encourages this group of pupils to be supported by their school to improve their decoding skills and the process evaluation found that the business-as-usual control pupils were experiencing a range of matched-time activities, including

differentiated literacy activities with the class teacher, small-group and one-to-one phonics activities. Therefore, it could be the case that both the intervention and control group made accelerated (or slower) progress in literacy compared to the rest of their class, but that GraphoGame was, on average, no more effective than the targeted literacy support put in place to support pupils with identified weak decoding skills.

Limitations

The most significant limitation of this evaluation is the moderate security of the impact estimate. Pupils found the pre-test difficult and, as a result, the mean and variance of test scores was low. This reduced the amount of statistical precision of the final impact analysis, increasing the minimum detectable effect size from 0.17 (as designed) to 0.25 (in reality). The trial had very low rates of attrition and the possible threats to validity (i.e. possible contamination between control and intervention pupils in the same class) were not observed during our school visits.

The trial was conducted in a small convenience sample of 15 schools in Cambridgeshire, so the results are not necessarily generalisable to a wider range of primary schools.

Future research and publications

This research sits within a wider debate on how best to teach reading in the early years of education. Since Sir Jim Rose's Review of the Teaching of Early Reading (Rose, 2006), schools in England have been required to support pupils to read through the use of systematic synthetic phonics. GraphoGame aims to teach pupils to read through the rhyme analogy approach based on 'rime'. Further research into the best methods for teaching early reading, and the circumstances and combinations in which different approaches to the teaching of phonics should be used, would be valuable for teachers and other practitioners.

References

Bhide, A., Power, A.J., & Goswami, U. (2013). A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter-based intervention. *Mind, Brain and Education,* **7** *(2*), 113-123.

Brem, S., Bach, S., Kucian, K., Guttorm, T.K., Martin, E., Lyytinene, H., Brandeis, D., and Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *Proceedings of the National Academy of Sciences, 107*(17), 7939-7944.

Department for Education (2016). *Statement from DfE about key stage 1 tests*. London: DfE [online]. Available: https://www.gov.uk/government/news/statement-from-dfe-about-key-stage-1-tests [15 November 2017].

Education Endowment Foundation (EEF) (2015). *Policy on analysis for EEF Evaluations*. London: EEF [online]. Available: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol/Analysis_for_EEF_evaluations_REVISED_Dec_2015.pdf [15 November, 2017].

Goswami, U. (1986) Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology,* **42**, 73-83

Goswami, U. (2005). 'Synthetic Phonics and Learning to Read: A Cross-language Perspective Usha Goswami', *Educational Psychology in Practice*, **21**, 4, 273–282.

Hedges, L. V., & Vevea, J. L. (1998). *Fixed- and Random-Effects Models in Meta-Analysis*. Psychological Methods, 3(4), 486-504. DOI: 10.1037/1082-989X.3.4.486

Howard-Jones, P. (2014). *Neuroscience and Education: A Review of Educational Interventions and Approaches Informed by Neuroscience Full Report and Executive Summary*. London: EEF [online]. Available: https://educationendowmentfoundation.org.uk/public/files/Publications/EEF_Lit_Review_NeuroscienceAndEducation.pdf [15 November, 2017].

Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically-motivated computer-assisted reading interventions, GG Rime and GG Phoneme. *Reading Research Quarterly,* **48** *(1),* 61-76.

Lovio, R., Halttunen, A., Lyytinen, H., Näätänen, R., & Kujala, T. (2012). Reading skill and neural processing accuracy improvement after a 3-hour intervention in preschoolers with difficulties in reading-related skills. *Brain Research, 1448*, 42-55.

Rose, J. (2006). *Independent Review of the Teaching of Early Reading*. Nottingham: DfES Publications [online]. Available: http://dera.ioe.ac.uk/5551/2/report.pdf [15 November, 2017].

Torgesen, J.K., Alexander, A.W., Wagner, R.K., Rashotte, C.A., Voeller, K.K.S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities. *Journal of Learning Disabilities*, **34** (1), 33–58. doi: 10.1177/002221940103400104

Worth, J., Sizmur, J., Ager, R. and Styles, B. (2015). *Improving Numeracy and Literacy: Evaluation Report and Executive Summary*. London: Education Endowment Foundation

Wyse, D. and Styles, M. (2007).'Synthetic phonics and the teaching of reading: the debate surrounding England's 'Rose Report', Literacy, **41**, 1, 35-42 [online]. Available: http://www.edalive.com/wp-content/uploads/2011/04/RoseEnquiryPhonicsPaperUKLA.pdf [15 November, 2017].

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:

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***			
5	Well conducted experimental design with appropriate analysis	MDES < 0.2	0-10%	5	Adjustment for Balance **[0]** Adjustment for threats to internal validity **[0]**	5
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%			
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%			
1	Comparison group with poor or no matching (E.g. volunteer versus others)	MDES < 0.6	51-50%			
0	No comparator	MDES > 0.6	>50%			

- **Initial padlock score:** lowest of the three ratings for design, power and attrition = 5 padlocks
- **Reason for adjustment for balance** (if made): none made, though there was imbalance at baseline for the analysed sample on both Phonics Screening (ES = -0.14) and NGRT (ES = -0.08) neither of these were significant, in addition prior attainment was taken into account (via the use of NGRT scores) in the analysis.
- **Reason for adjustment for threats to validity** (if made): none made
- **Final padlock score:** initial score adjusted for balance and internal validity = 5 padlocks

Appendix C: Memorandum of Understanding

Neurocognitive Factors Governing Response to Intervention with GraphoGame Rime

Dear [name of headteacher],

We are very pleased and grateful that your school wishes to take part in our research study "Neurocognitive Factors Governing Response to Intervention with GraphoGame Rime".

This document will help to clarify what is involved and act as a memorandum of understanding between your school and the researchers involved.

About the research teams

Two teams of researchers will carry out the evaluation. One of these teams, which is based across the Centre for Neuroscience in Education and the Department of Education at the University of Cambridge, has developed an English version of a software game for teaching phonics knowledge via rhyming patterns, called GraphoGame Rime. The English team developed the game from an original game developed by a Finnish University. The Finnish game has been very effective and is now in every Finnish primary school. The Cambridge team will provide training and support to teachers on how to deliver the English game, GraphoGame Rime, within the classroom. To identify which factors predict how much benefit children from the game, they will carry out tests (pre-tests) of the eligible children's reading skills along with other measures, such as memory and attention, at the start of each year of the project.

The other team, who work at the National Foundation for Educational Research (NFER), will be responsible for evaluating the effectiveness of GraphoGame Rime. This team will test the eligible children's reading and spelling skills at the end of each year of the project (post-tests) to find out whether the children's skills were affected by use of GraphoGame Rime.

About the research

The research teams will hold a meeting with representatives from all schools involved in the evaluation to clarify what is planned. The evaluation will initially involve selecting Year 2 pupils who do not meet a predetermined standard in the Year 1 phonics check. Parents of these pupils will be given the opportunity to opt out of the trial. Pupils will then be allocated at random into two groups: a group of pupils that spend 10-15 minutes each day playing the GraphoGame Rime computer games during literacy lessons and a 'business as usual' control group of pupils from the same classes. The number of pupils that will be in the intervention and control groups respectively will be known by teachers long in advance, while the identity of the pupils that are allocated to each group will not be revealed until after the pre-testing has been completed.

The evaluation will consist of a pre-testing phase in September-December 2015, the implementation of GraphoGame Rime in January-April 2016 and post-testing in May-June 2016. The same sequence of events will repeat in the following academic year with a new cohort of Year 2s.

Data sharing[9]

All the information that we gather about individual pupils, teachers and schools will be kept completely confidential in accordance with the Data Protection Act. No information about individual children will be

[9] Applies to pupils whose parents have not opted out of the study. It should be noted that all parents have the right to withdraw their child from the study at any time.

made available to anyone outside of the research teams within the University of Cambridge, NFER, Education Endowment Foundation (who fund the work) and the Fischer Family Trust (who will collate and anonymise the data for upload to the UK Data Archive). We will not use pupils' or teachers' names or the names of any of the schools involved in the project when we write about the results of the research. Our accounts of the effectiveness of GraphoGame Rime will be presented in the form of aggregated or averaged data.

The NFER team will be requesting the following information about the Year 2 pupils in your school: Unique Pupil Number (UPN), name, date of birth, score on the phonics screening check, eligibility for free school meals and the class number/name for each pupil. This information will be used to determine which pupils are eligible for the project, for analysis and for selecting an even number of pupils in each class. The information will be requested at the start of each year of the project.

The NFER team also plans to make a request for further information about the children in the project that is held in the National Pupil Database (NPD). The team will make a request for Unique Pupil Numbers (UPNs) which will make it possible for them to access data about individual children in the NPD.

Process evaluation

As part of its evaluation, the NFER team would very much like one of its members to visit half of the schools involved to understand how GraphoGame Rime is being used in practice. The visits would take place in both years of the project and would involve observation of GraphoGame Rime in use, interviews with staff involved in the project and if possible, small group interviews with pupils. The NFER team would also like to gather the views of the other half of schools through telephone interviews.

We do hope that you will be happy about these plans. If you have any concerns about them, please contact us. You can telephone or email [contact details].

Once again, we would like to express our gratitude to you for joining us in this research. We believe that it will be a thoroughly worthwhile project and that it will produce some valuable results.

Summary of responsibilities

The University of Cambridge will:

- Deliver an introductory session and a training session for teaching staff
- Provide access for the school to the Graphogame software
- Be the first point of contact for any questions about the evaluation
- 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 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 pupils are allocated to treatment or control)
- Allow time for each testing phase and liaise with the evaluation team to find appropriate dates and times for testing to take place
- Release staff so that they can attend the introductory session and the training session
- Ensure the shared understanding and support of all school staff for to 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 GraphoGame Rime as detailed above:

Headteacher: _____

School name: _____

Date: _____

Professor Usha Goswami, University of Cambridge

[INSERT SIGNATURE]

Dr Ben Styles, National Foundation for Educational Research

[INSERT SIGNATURE]

Appendix D: Parent Letter and Information Sheet

Neurocognitive Factors Governing Response to Intervention with GraphoGame Rime

Dear Parent/Guardian,

My name is Usha Goswami and I am Director of the Centre for Neuroscience in Education at the University of Cambridge. We recently secured funding from the Education Endowment Foundation and the Wellcome Trust for a project which will evaluate the educational impact of 'GraphoGame Rime' for Year 2 pupils. GraphoGame Rime is a computer program which we designed to help children to improve their reading skills through interactive activities based on rhyme. Our aim is to make a research-based game that will be free to the end-user (see info.graphogame.com)

This project is running across several schools in Cambridgeshire as well as here at [school name] and I am writing to you to make you aware of what is involved in the project. Please feel free to contact me by email or phone if you have any concerns. I can be contacted via email on [email address] or by telephone on [telephone number].

I have attached an information sheet which explains in simple terms what is involved in agreeing for your child to participate. This is a really good opportunity to help improve reading and we hope that as many pupils as possible will be able to participate. The headteacher in your child's school has given us general permission to conduct the study in the school, but we also want to offer you the chance for your child to opt out of the project, if you so wish.

Please return the reply slip at the bottom of this letter to your child's teacher if you wish to opt out of the project. If we do not hear from you by [date] we will assume that you have no objections and your child will be asked to take part.

Yours faithfully,

Graphogame Research Project

I do not want my child to be asked to participate in this project.

Child's Name: _____

School Name: ----------

Parent's Name and Signature:

_____

Parent Information Sheet

Study title
Neurocognitive Factors Governing Response to Intervention with GraphoGame Rime

What is the purpose of the study?
We wish to evaluate the impact of using GraphoGame Rime on children's reading attainment. GraphoGame is a computer game designed by us to help children to understand the building blocks of reading, and uses words in rhyme families.

Why have we been approached?
Your child has been approached because we hope to recruit Year 2 pupils who have the potential to achieve more in reading. In particular, we are interested in children whose reading achievement could be strengthened by practising letter-sound correspondences to enable them to engage more successfully with the primary school curriculum ahead of them.

Do we have to take part?
No, participation is entirely voluntary and there are no consequences if you or your child decides not to take part.

What will happen to my child if s/he takes part?
If your child takes part, they will be randomly selected to experience either 'GraphoGame' or normal classroom activities during the Spring term of the school year. GraphoGame involves the children spending 10-15 minutes of their usual literacy lessons developing their reading skills through the interactive computer-based activities. Each child has a personal log-in and the game offers increasingly challenging levels as they improve their skills. S/he will also be assessed for attention, cognitive and sound awareness skills in 3 sessions of 20 minutes each in the Autumn term, to explore which skills predict who benefits most from playing GraphoGame. S/he will be assessed on his/her reading and spelling in the Autumn term and again at the beginning of the Summer term.

What are the possible disadvantages and risks of taking part?
There are few disadvantages and risks. The reading and spelling assessments do take a little time to complete but we will ensure that they are completed at a time when it will cause minimal disruption to your child's school work. The assessments may feel challenging and some children may feel a little self-conscious about completing them (for example, if their reading is not as good as they would like it to be). However, all results will remain confidential to the research team and we will do our best to put your child at ease throughout. The results will also be shared with your child's teacher to help them plan the best ways to continue to improve your child's reading skills.

What are the possible benefits of taking part?
We know from our previous work developing GraphoGame that using GraphoGame helps children's spelling and their ability to break down words into sounds. We do not know whether this translates into improved reading at school. If we find that it does, then we anticipate that more schools will offer GraphoGame as a reading support method within the school day.

What if something goes wrong?
You can indicate to the teacher if you no longer wish for your child to take part, and you can leave the study without question. If you are unhappy with the conduct of the study, you can contact me directly in the first instance using the number at the end of this sheet. You are free to withdraw at any point during the study, and for up to one month following the completion of the study. You can do this by contacting me and giving me your child's name and the name of his/ her school. If you are unhappy with the

conduct of the research team, you can contact me (Professor Usha Goswami) in the first instance, or Professor [Name], who is chair of the University Ethics Committee (address: [address]).

Will my taking part in this study be kept confidential?
Yes. Pupils' test responses and any other pupil data will be treated with the strictest confidence. We will not use your child's name or the name of the school in any report arising from the research. All cognitive data will be identified only by a code, with personal details kept in a locked file or secure computer with access only by the immediate research team. The reading test responses will be collected by GL Assessment and accessed by researchers at the National Foundation for Educational Research (NFER). Reading and spelling test sheets will be stored in a locked cabinet and destroyed as soon as the scores have been saved in a computer file. For the purpose of research, the responses will be linked with information about your child from the National Pupil Database (held by the Department for Education) and shared with the University of Cambridge, NFER, Education Endowment Foundation (EEF), EEF's data contractor FFT Education and, in an anonymised form, the UK Data Archive.

What will happen to the results of the research study?
The reading test data will be used as the basis of a report to be compiled by colleagues from the National Foundation for Educational Research, which will be submitted to the Education Endowment Foundation. Reports based on the cognitive/attention scores, reading scores and spelling scores will be presented at academic conferences and it may also be written up for publication in peer-reviewed academic journals. Crucially, the Education Endowment Foundation report will be published and freely accessible, so participating schools and families will be informed as to the outcomes of the project overall.

Who is organising and funding the research?
The research is organised by Professor Usha Goswami, who is Director of the Centre for Neuroscience in Education at the University of Cambridge and the developer of GraphoGame Rime. The research is funded by the Education Endowment Foundation.

Who has reviewed the study?
The Education Endowment Foundation and the University of Cambridge Psychology Ethics Committee have reviewed and approved this study.

Contact for further information
We expect that your child will enjoy doing the tests and being part of the programme. Your child may withdraw at any time. If you prefer for your child NOT to take part, please inform their teacher. If you would like more information, please contact:

Professor Usha Goswami
Address: [address]
Tel: [telephone number]
Email: [email address]

Appendix E: Randomisation Syntax

```
Title 'Pupil randomisation'.
get file = "I:/EERR/First year/Definitive list of eligible pupils.sav".

* Set randomisation seed.
set rng=mt, mtindex=08122015.

* Randomise school order.
compute schrand=rv.uniform(0,1).
sort cases by NFERNO.
aggregate outfile = "c:/temp/schools.sav" /break = NFERNO
  /sch = first(schrand).
match files /file = * /table = "c:/temp/schools.sav" /by NFERNO.

* Randomise class order.
compute clsrand=rv.uniform(0,1).
sort cases by NFERNO Class.
aggregate outfile = "c:/temp/class.sav" /break = NFERNO Class
  /cls = first(clsrand).
match files /file = * /table = "c:/temp/class.sav" /by NFERNO Class.

* Create variable for randomising pupils.
compute rand=rv.uniform(0,1).

* Stratify randomisation by class.
sort cases by sch cls rand.

* Allocate pupils to groups by order.
compute group = MOD($casenum,2).
variable labels group = "Group allocation".
value labels group 0 'Control group' 1 'Intervention group'.

* Create string variable for exporting in Excel.
string groupallocation (A20).
if group = 0 groupallocation = 'Control group'.
if group = 1 groupallocation = 'Intervention group'.

* Group split overall.
frequencies group.

* Group split by school/class.
crosstabs NFERNO by group.
crosstabs Class by group by NFERNO.

* Sort data by group so that Excel file is grouped.
sort cases by Class group NFER_PUPIL_ID.

* List of schools and codes.
frequencies NFERNO.

sort cases by NFER_PUPIL_ID.

save outfile = "I:/EERR/First year/Randomisation.sav".
output save outfile = "K:/EERR/Analysis/First year/Randomisation.spv".
```

Appendix F: Process evaluation achieved sample

School ID	Year of trial	Visit/ telephone	Teacher interview/s (or interview with other person supervising GraphoGame)	Senior leader interview/s	Observation of pupils playing GraphoGame	Pupil interview/s (No. of pupils)
A	1	Visit	3	-	1	1 (5)
	2	Visit	3	-	1	1 (2)
B	1	Visit	2	1	1	1 (3)
	2	Visit	2	-	1	1 (2)
C	1	Visit	2	1	1	1 (3)
	2	Visit	1	1	1	1 (3)
D	1	Visit	1	1	-	1 (3)
	2	Not in trial	Not in trial	Not in trial	Not in trial	Not in trial
E	1	Visit	1	1	1	1 (3)
	2	-	-	-	-	-
F	1	Telephone	1	-	-	-
	2	-	-	-	-	-
G	1	Telephone	-	1	-	-
	2	Telephone	1	-	-	-
H	1	Telephone	1	-	-	-
	2	Telephone	1	-	-	-
I	1	Telephone	1	-	-	-
	2	Telephone	1	-	-	-
J	1	-	-	-	-	-
	2	-	-	-	-	-
K	1	Telephone	1	-	-	-
	2	-	-	-	-	-
L	1	Not in trial	Not in trial	Not in trial	Not in trial	Not in trial
	2	Visit	1	-	1	1 (3)
M	1	Not in trial	Not in trial	Not in trial	Not in trial	Not in trial
	2	Telephone	1	-	-	-
N	1	Not in trial	Not in trial	Not in trial	Not in trial	Not in trial
	2	Telephone	1	-	-	-
O	1	Not in trial	Not in trial	Not in trial	Not in trial	Not in trial
	2	Visit	2	1	1	1 (2)
TOTAL			**27**	**7**	**9**	**10 (29)**

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 v3.0.

This information is licensed under the Open Government Licence v3.0. To view this licence, visit **http://www.nationalarchives.gov.uk/doc/open-government-licence/**

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**

Education
Endowment
Foundation

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