



Education  
Endowment  
Foundation

# Fresh Start

Evaluation Report and Executive Summary

February 2015

## **Independent evaluators:**

Stephen Gorard

Nadia Siddiqui

Beng Huat See



**Durham**  
University

## The Education Endowment Foundation (EEF)



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

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;
- 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-The Private Equity Foundation) and received a founding £125m grant from the Department for Education.

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



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

**James Richardson**

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

**p:** 02078021924

**e:** james.richardson@eefoundation.org.uk

**w:** www.educationendowmentfoundation.org.uk

## About the evaluator

The project was independently evaluated by a team from Durham University led by Professor Stephen Gorard.

Stephen Gorard is Professor of Education and Well-being, and Fellow of the Wolfson Research Institute at Durham University. He is a Methods Expert for the US government Institute of Education Science, member of the ESRC Grants Awarding Panel, and Academician of the Academy of Social Sciences.

His work concerns the robust evaluation of education as a lifelong process, focused on issues of equity and effectiveness. He regularly advises governments and other policy-makers, including oral and written advice to the House of Commons Education Committee every year since 2003. He is also a widely read and cited methodologist, involved in international and regional capacity-building activities, and used regularly as an adviser on the design of evaluations by central and local governments, NGOs and charities. He is currently an evaluator for the European Commission Directorate-General for Education and Culture, the Department of Work and Pensions, the Food Standards Agency, the Learning and Skills Information Service, and the Education Endowment Foundation. He is author of nearly 1,000 books and papers.

### Contact details:

#### **Professor Stephen Gorard**

School of Education  
Durham University  
Leazes Road  
DH1 1TA

**p:** 0191 334 8419

**e:** [s.a.c.gorard@durham.ac.uk](mailto:s.a.c.gorard@durham.ac.uk)

---

## Contents

<b>Executive summary.....</b>	<b>4</b>
<b>Introduction .....</b>	<b>6</b>
<b>Methods .....</b>	<b>10</b>
<b>Impact evaluation results .....</b>	<b>17</b>
<b>Process evaluation results.....</b>	<b>25</b>
<b>Outcomes.....</b>	<b>27</b>
<b>Conclusions and implications .....</b>	<b>30</b>
<b>References.....</b>	<b>32</b>
<b>Appendices.....</b>	<b>34</b>

## Executive summary

### The project

Fresh Start (FS) is a catch-up literacy intervention for pupils at risk of falling behind their peers in early secondary schooling. It provides systematic and rigorous practice in phonics so that pupils are at an appropriate level to join the mainstream group after completion of the intervention. Pupils are assessed and then grouped according to their levels of reading ability. Teaching in these groups begins with recognition, practice and blending of sounds and graphemes, based on a set of module booklets.

This evaluation involved 433 pupils in Year 7, in ten secondary schools, who had failed to achieve a 'secure' National Curriculum Level 4 (4b and above) in their primary KS2 results for English. Of these, 212 within all schools were randomly allocated to receive the intervention for 22 weeks during their first year at secondary school. This shorter time period was partly justified because the pupils are at secondary school. The other 221 pupils were randomly allocated to a control group in phase one, and received the intervention the following year. The intervention was organised by taking pupils out of regular English lessons for one hour, three times per week.

### Key conclusions

1. Fresh Start shows considerable promise as an effective catch-up intervention for low-attaining readers at the transition phase from primary to secondary school.
2. Pupils must be grouped according to the reading scores obtained from a Fresh Start screening test. Each group has a homogeneous level in terms of pupil reading abilities
3. Continuous feedback to teachers and support from trainers assists successful implementation.
4. It is feasible for schools to conduct evaluations of their own planned interventions, under favourable circumstances, and with some advice and oversight from independent expert evaluators.

This was a different kind of study to the trials usually funded by the EEF. In 2013, three secondary school clusters, consisting of 10 schools in total, applied independently to EEF for funding to set up a programme for FS, and simultaneously evaluate its impact in their own schools. Each application was deemed too small in scale to run a successful evaluation of the programme, but if the schools involved were to cooperate then the scale would be sufficient for an 'aggregated' efficacy trial. Efficacy trials seek to test evaluations in the best possible conditions to see if they hold promise.

Each school ran a small trial of FS in isolation, and made all of the relevant evaluation decisions such as randomly allocating pupils to the control or intervention group. Researchers at Durham University were assigned as independent evaluators for this trial. Their roles were to advise the school leads on the process of conducting research, randomisation and testing, and to aggregate the eventual results from all schools.

There was no direct involvement from FS in the initial funding proposal to the EEF, but the developers were used by schools to provide training.

### What impact did it have?

The FS pupils in the intervention group made more progress in literacy than the control group after 22 weeks. The overall effect size was +0.24 in terms of the gain between the scores in the test before the intervention (the pre-test) and the test after the intervention (the post-test). This is equivalent to approximately 3 months of additional progress in reading age. Progress was assessed as the gain in scores between a pre-test (GL Assessment's New Group Reading Test A) and a post-test (New

Group Reading Test B). The impact evaluation also suggests positive progress results for FSM-eligible pupils, and for all pupil sub-groups regardless of age, sex, first language, ethnicity, or special education need. However, as the numbers of pupils in the sub-groups are smaller, these findings are less secure than the overall finding.

The intervention was observed by the evaluation team as being generally well conducted, and attractive to teachers and pupils. The intervention was well received by the schools. All participating schools have plans to continue Fresh Start for future cohorts.

### How secure is this finding?

The existing literature suggests considerable promise from using phonics in general, and Fresh Start more specifically. However, the impact of FS has not been demonstrated in the UK, and so this efficacy trial is the first of its kind in England.

The primary outcome measure was reading comprehension as measured by an independent test. The results were analysed with gains in overall reading scores as well as with standardised age scores, and both gave similar effect sizes. A pre-test indicated there was considerable imbalance between the groups (with the treatment group having lower prior attainment). Attrition was very low only 2%. Only 14 pupils have missing test scores. The control group activity was 'business as usual' and these pupils were not given access to the programme. The intervention was regularly administered, the process was closely monitored and there was no indication of pupils or teachers becoming demoralised after the randomisation occurred. Therefore, the evidence achieved can be interpreted as having a security rating of three padlocks.

Participation in the Fresh Start intervention was at the instigation of the school leaders and cluster heads. They were already enthusiastic about the programme. This may limit the generalisability of results to other schools who may be less enthusiastic about its potential and therefore not deliver the intervention as effectively.

The process demonstrated that schools are capable of running trials, but only with appropriate expert help. The main area of concern was in allowing schools to do the randomisation. Following the randomisation process, there was a noted imbalance in the pupils' pre-test scores in the control and intervention groups. A number of statistical techniques were applied to control for the imbalance in the two groups in the final analysis. The results suggest that the outcomes of the trials are unlikely to be caused by bias or dropout of the 14 cases.

### How much does it cost?

The direct cost to schools is for the Fresh Start training and material resources required in the intervention. For each pupil the cost of the material is £95.90. The cost of staffing to assist teachers depends largely on each school's initial position. So, for example, a school with four relevant teachers to be trained and 20 pupils per year group, would have to pay £2,316 or around £116 per pupil.

Group	No. of pupils	Effect size	Estimated months' progress	Evidence strength	Cost rating
<b>Intervention vs control (all pupil)</b>	419 (10 schools)	+0.24	+3 months		££
<b>Intervention vs control (FSM)</b>	104	+0.24	+3 months		££

*\*For more information about evidence ratings, see Appendix 1 in the main evaluation report. Evidence ratings are not provided for sub-group analyses, which will always be less secure than overall findings.*

*\*\*For more information about cost ratings, see Appendix 2 in the main evaluation report.*

## Introduction

### Intervention

The Fresh Start (FS) literacy intervention is one of the four Read Write Inc. literacy programmes developed by the Ruth Miskin Foundation. FS is a 33-week literacy intervention for struggling readers aged 9 to 13. The intervention has its own screening assessment criteria to determine pupils' initial reading levels and guide their allocation to small class groups for FS lessons (this is explained below). The FS teaching approach is based on knowledge and practice of phonics, and word recognition according to pure sounds. FS is a complete literacy catch-up package that targets improvements in both reading and writing, but in this randomised control trial the impact has been assessed only for reading comprehension, and after 22 weeks. This was the decision of the school leaders who wanted to assess the impact on reading first. The length of the intervention was reduced to 22 weeks to see the impact for one complete term between pre- and post-test. The schools took this decision in order to have least disruption in regular school activities.

The underlying premise of FS is that language comprehension and word recognition are both essential skills for becoming successful readers. In order to overcome the problems of reading comprehension faced by some pupils, the main challenge at the phonological and grapheme levels needs to be addressed in the initial stage of the learning process. The pupils' main challenge is to understand the relationship between sounds and written letters. Reading comprehension follows once pupils understand the principle of phonics and recognise the meanings of words. FS comprises 33 modules along with phonics teaching plans, sound cards sets and a sound chart, practice exercises, worksheets and assessment exercises for pupils. FS teaching also involves specific strategies for modelling sounds, encouraging practice among pupils, motivation and praise.

### Background

According to the National Curriculum Assessment (2012-2013) about 25 per cent of pupils leave primary schools without achieving the expected Level 4 in English, and there are nearly 767 mainstream primary schools (6 per cent of the total number of schools in the UK) that are below expected primary school floor standards (DfE 2013). Attainment in reading has been an important focus for the Department for Education in England, and there have been various policies implemented in order to achieve the national targets of attainment in reading at Key Stage 2. More generally, the government introduced the scheme of pupil-premium funding for schools to raise the attainment level of disadvantaged pupils. Schools now have the opportunity to use these funds to build capacity and buy resources or teaching approaches that can support pupils who are at risk of falling behind. An essential step is to identify teaching approaches that are effective for pupils at risk, and that are also practical for integration within the National Curriculum and the syllabus used in the schools.

Fresh Start literacy programmes are already widely used interventions to address the literacy problems of disadvantaged pupils in primary schools. The FS approach is commonly known as a 'systematic synthetic approach' using 44 basic sounds in English and sounding out the letters of the printed words followed by blending the sounds to form a complete word (definition of systematic synthetic approach from Torgerson 2003). The sounds, identified with printed letters, are taught first in isolation and children later learn how to blend and synthesise the segmented sounds to read the whole word. Writing is the reverse of this sequence where every sound of the word is first uttered and then children write respective letters for each of the sounds (Brooks 2003).

The approach of teaching reading through phonics gained popularity in England in 2006 when the Rose Report (2006) presented the findings of a review of the National Literacy Strategy (NLS) implemented in 1998. The NLS involved an 'analytic approach' in which the 26 letters are taught first and then words are introduced to recognise sounds associated with each letter. Children learnt to

deduce the sound of each letter in the context of the whole word (Johnston and Watson 2005). The Rose Report reported no evidence that the NLS reforms achieved the set targets in literacy, and recommended changes in reading pedagogy.

Debates about the effectiveness of different teaching approaches have affected nationwide policies on literacy and education attainment for decades (Teaching Children to Read: Eighth report of session 2004-05: <http://www.publications.parliament.uk/pa/cm200405/cmselect/cmmeduski/121/121.pdf>). However, there is a lack of solid evidence about which teaching approaches are most effective and under what conditions these approaches work or fail to work. One study in Clackmannanshire involving 92 Year 1 pupils (Johnston and Watson 2005) claimed positive effects for the synthetic phonics approach. This study compared the results of teaching children using synthetic phonics and the whole-word approach. Children were pre-tested and divided into three groups: two groups were given a synthetic phonics intervention and an analytic phonic intervention respectively, and the third group was a control group that continued with their usual English lessons. The researchers administered the interventions to each group for 10 weeks (twice a week, 20 minutes a day). The two groups were taught 114 target printed words by the synthetic phonics and whole-word approaches. Results of standardised tests showed that pupils taught through the synthetic phonics approach showed more progress than those who were taught through whole-word approach and those who continued normal English lessons in the classes. This study has been criticised for its process of implementation where the different interventions did not standardise the rules of introducing and practising new sounds and letters across the three groups (Wyse and Goswami 2008). It was also a small study with only about 30 pupils per group.

Another study conducted in Clackmannanshire reported further evidence in support of teaching reading through phonics (Johnston and Watson 2004). The sample included 304 pupils in Year 1 from 13 primary schools. The target sample was divided into three groups, each group being given a different literacy intervention targeting pupils' word knowledge, spellings and reading comprehension. The authors claimed that matched group sampling was not possible therefore the researchers decided to include the most disadvantaged pupils in a group that was taught through the synthetic phonics approach (p.12). This group imbalance from the outset may explain why pupils taught through synthetic phonics seemed to have made faster progress than pupils in the other sub groups. It could be regression to the mean. The report, however, does not give detailed information on the seven-year-long follow-up results of the initial target pupils.

In the United States there have been similar debates about the effectiveness of reading pedagogy. In 2000, the National Reading Panel report (<http://www.learningpt.org/pdfs/literacy/nationalreading.pdf>) claimed to have found evidence of the positive impact of systematic phonemic awareness approaches for primary school children. A What Works Clearinghouse (WWCH 2010) review of four randomised controlled trials on 'Sound Partners', which is very similar in concept and pedagogy to FS, showed a medium to large positive impact of 'Sound Partners' on letter recognition, fluency and comprehension, but no effect on reading achievement of beginning readers. The evidence is thus mixed, with this kind of intervention appearing to be effective for some measures but not for others.

A meta-analysis conducted by Torgerson et al. (2006) of 12 relevant randomised controlled trials found that systematic phonics teaching approaches (teaching of letter-sound correspondences in an organised, regular, explicit and sequenced manner, Ehri et al. 2001) are effective for all pupils in comparison to unsystematic or no-phonics instructional approaches. The cumulative effect size reported for the systematic phonics approach was 0.27. The review also stated that due to the poor quality of evidence, very low number of experimental studies and lack of clear reporting on sampling and method of random allocation, it is very difficult to conclude on the effectiveness of any one of the systematic teaching approaches on reading comprehension and spellings.



FS has also been shown to have mixed effects for secondary pupils. A systematic review which looked at 14 reading interventions for secondary school pupils showed that FS (three relevant studies) had a positive effect on reading comprehension, and the effect sizes published were 0.34 and 0.25, but no benefit was found for spelling accuracy (Brooks 2007). The total sample size in each of the respective studies was only 29, 62 and 156. In each of the three studies there were problems of missing data, lack of comparison group and poor reporting of group allocation. It is also not clear in any of the studies if the sample had established baseline equivalence and also if treatment groups were randomly allocated. The overall evidence is therefore weak.

In the UK, a group of schools in the local authority of Leicester adopted FS for secondary school pupils who were struggling to meet the literacy targets (Lanes et al. 2009). A pre- and post-test research design was employed to measure the impact of FS. The study included 63 pupils who were three years behind the chronological reading age and had severe reading difficulties. The evaluation report available for this programme does not contain vital information about the sampling and the test data. It is therefore difficult to assess the effectiveness of the FS intervention. There is substantial process evaluation data available that suggests that the pupils on the programme showed improvement and the teaching staff were very positive about the FS method and resources. This could be misleading in the absence of impact evidence.

An evaluation study commissioned by the DfE to evaluate the effectiveness of Fresh Start for disadvantaged pupils at the Key Stage 3 level also did not yield conclusive results (Brooks et al. 2003). The study reported problems with attrition and testing. As a result only 30% of the 500 pupils in the sample were included in the final analysis. The positive impact reported on the progress in reading comprehension is thus not really warranted. Dropouts are not random, so using the results of those who remained would not give an accurate picture of effectiveness.

The existing evidence available on FS suggests that the intervention is popular among teachers. Many gave positive feedback. Ofsted (2010) mentioned Read Write Inc. literacy programmes as one of the interventions adopted by the 'best' schools in the UK. However, the evidence so far has been mixed. FS was found to be effective on some measures of literacy, but not for others, and in some studies. In each case the rigour of studies has been challenged by scale, non-random sampling, attrition and missing data. No study provided a cost-effectiveness analysis of FS. This is crucial information for schools planning to adopt the approach in their classroom teaching.

## Objectives

- To evaluate the impact of using FS for 22 weeks on the literacy performance of Year 7 pupils who had not previously secured Level 4b or above in KS2 English
- To evaluate the impact of FS for pupils from disadvantaged backgrounds
- To record the fidelity of the intervention through process evaluation and assess its feasibility at a larger scale
- To assess the extent to which schools themselves can conduct RCTs, with advice and guidance from expert evaluators.

## Project team

Fresh Start is a commercially available and specifically tailored literacy programme offered by Read Write Inc. In this evaluation the developers of FS had no direct involvement in the design of the study. However, FS training and school support visits have been conducted by the FS trainers (which is standard practice). The intervention was conducted independently by the volunteering schools and the evaluation process was supervised by the independent evaluation team. Ten secondary schools participated in the trial, implemented the FS intervention and conducted the randomised controlled trial RCT with advice from the independent evaluators. A big thank you goes to all the schools,

education institutions and teachers involved in this research, who ran the projects within their own schools and contributed a great deal to the evaluation. Thanks to the Harlow Education Consortium, Withernsea High School, South Holderness Technology College and Hornsea School and Language College, who indicated they were happy to be named in the report.

The evaluators from Durham University provided training in evaluation for the school leads, conducted a process evaluation, and led the impact evaluation.

## Methods

### Trial design

This evaluation project was a two-arm and school-led randomised controlled trial involving 10 schools from three clusters or federations (with 5, 3 and 2 schools each). The target group of 433 pupils was selected based on KS2 scores; pupils at Level 4c and below in English were selected, and analysed using intention to treat. The intervention was carried out from September 2013 to February 2014.

After selection of the target group, the schools conducted the pre-test and then each cluster randomised the target group individually into two arms. The treatment group received FS and the pupils in the control group continued their regular English lessons as per normal during this phase (some of which also involved working in small groups). FS is a very specific and unusual approach to literacy which must be applied as a whole, and was conducted by one or more specially trained members of staff in each school. Therefore the threat of diffusion exists but is judged as minimal.

It was a waiting list design in which pupils in the control group received FS after the post-test was completed in February 2014. This design is ethical since no pupil is denied the intervention and does not demoralise pupils post-allocation (Gorard 2013). The schools were keen to be involved so pupil-randomisation at the school level kept each of the 10 schools engaged in the intervention. This design minimised the possibility of schools dropping out of the study.

The three school cluster leads managed the interventions and the trial in their respective schools. The evaluators had direct access to each school in the trial but having a cluster-lead in each of the three clusters made communication with the evaluators simpler. The cluster-leads were school managers or school partnership heads in their local authority.

The developers of FS were not directly involved, which gave flexibility to the cluster heads and school leaders to manage the intervention according to the available space, time and the needs of the schools.

As this is a school-led trial the school leaders sought parental consent of their children's involvement in the intervention and trial that involves testing. The schools used their usual communication strategies such as letter or email to parents/guardians for their opt-out consent. This procedure was followed by the school leaders before randomisation.

Ethical approval for the role of the independent evaluators came through the Durham University Ethics Committee.

### Testing

The New Group Reading Test was developed by GL assessment and the National Foundation for Education Research. NGRT has A and B versions of the test and is ideal for a pre- and post-test design. The age-appropriate level NGRT suitable for the age group 10-13 (Year 5 to Year 8) was selected (<http://www.gl-assessment.co.uk/products/new-group-reading-test/test-detail>). In both the A and B versions of the test the items are 20 sentence completions and 4 short passages for context comprehension and reading comprehension. The areas of assessment are as follows:

## Vocabulary

- Grammatical knowledge
- Inference skills
- Ability to recognise
- Authorial intent
- Deduction skills

The digital version of the test was used for this evaluation. There is no time limit for completion of the test. However, it is estimated that the test would take no more than 45 minutes to complete. The test difficulty level adapts according to pupil's initial responses. Not all pupils in a group received the same question items which reduces the chances of cheating among pupils. The test was agreed by all parties before the trial began in September 2013.

## Eligibility

**Schools:** There were three clusters, altogether having 10 schools, which proposed implementing FS intervention as a literacy catch-up programme for their Year 7 pupils. This was considered as an opportunity by the EEF to run a school-led aggregated trial. Six of these schools had already used the FS programme but needed to build more capacity and resources. However, the other four schools implemented FS the first time. The schools participating in this evaluation were responsible for running the RCT with close guidance and advice from the evaluator.

**Pupils:** The intervention was targeted at pupils aged between 12 and 13 (Year 7) at Key Stage 3 with Level 4c and below in KS2 English. The overall target group had a considerable number of pupils eligible for FSM, and with learning difficulties and special education needs (SEN). Parental consent for their child to opt out from the trial was sought by the schools. The opt-out consent letters were sent only to those pupils who were selected in the target group.

## Intervention

FS is a literacy intervention in which reading is taught through awareness and practice of phonics and then blending word segments to form complete words. The intervention involves systematic teaching of 44 sounds in English and practice of blending the sounds until pupils start recognising and reading the words. FS aims to support young pupils who somehow missed earlier opportunities and FS is an extra chance to fill those learning gaps. Therefore, FS is tailored to get the pupils off the intervention quickly so that they can participate in mainstream literacy activities.

The learning process begins from pure sounds, and then letters are introduced as graphemes representing each sound. FS is a catch-up literacy package with modules to be completed in 33 weeks. The modules are graded according to the reading age level. The programme recommends an FS session every day instead of the usual English lessons and each FS session should be one hour long. However, for the current research the dosage of FS has been reduced to three weekly hour-long sessions for 22 weeks. The schools decided to assess the impact of FS on reading comprehension for one complete term, and wanted to minimise disruption in the school schedules from the research activity.

The package of materials includes assessment charts, lesson plans for each module and extra resources for teaching. The FS intervention has the following main features:

- Starting point and pupil grouping
- Phonic lessons
- Assessment and modules
- Teaching strategies and classroom management style

- Starting point and pupil grouping

The intervention begins with assessment of pupil's phonics and simple word recognition. The tests are included in the package and the teacher conducts the test on individual pupils. The tests require pupils to read aloud or sound out the letters and words to the teacher. A pupil is asked by the teacher to read the written word on sight or to first break the written words into sound segments and then blend. The teacher marks each word on the basis of pupil's knowledge of the sound of letters and speed of reading.

It is suggested in the FS teacher's guide that after the screening test pupils can be divided into four ability groups according to the scores achieved in the first assessment test. FS claims that the closer the homogeneity within the groups the faster will be the individual pupil's progress. FS emphasises this homogeneity of pupils' reading ability within the groups more than the group size. However, small groups are recommended for effective results. FS recommends that the teacher also provides 20-minute regular one-to-one FS sessions, depending on pupils' individual needs.

### Phonic lessons

After grouping, pupils are introduced to phonics and the corresponding letters. The phonic lessons are systematic teaching of 44 sounds that are divided into three sets of Speed Sounds. Letters are introduced as graphemes and explained as written forms of sounds. This stage mainly focuses on pupils' decoding ability. The teacher introduces the concept of sound and grapheme with the help of a chart and Speed Sound cards. This stage begins with teaching and practice of Simple Speed Sounds of Set 1 and Set 2. According to the National Curriculum reading levels, the sounds and graphemes in Sets 1 and 2 can be read by pupils from 1b to 2b Levels.

Teachers also introduce blending sounds in simple words through Sound Talk (sounding out) and pupils repeat the sound after the teacher. Speed Sounds and blending sounds at this stage are assisted by different mnemonics (picture cards, picture books, Fred puppet, talking fingers). Pupils practise a set of words written on the cards. The teacher then comes to a level where pupils can do independent blending of sounds. A set of nonsense words is used to practise independent blending of sounds eg. rud, vun, chab, bruck.

Writing letters is also taught and practised but letters are never mentioned by their names. Three steps of practice followed throughout this stage are 'say it', 'read it' and 'write it'. This stage of Speed Sounds practice is three weeks long and then the series of modules are introduced. Speed Sounds Set 3 has complex sounds and graphemes which are introduced at module level 4. This set of sounds and graphemes can be read by pupils between NC reading Levels of 2b to 2a and above.

### Assessment and modules

Pupils in the third week take the entry test in the FS package. The pupils are individually tested by the teacher. At that stage pupils start being withdrawn from other English literacy lessons and grouped to attend FS sessions. The entry test includes six passages selected from the modules. Each passage is a sample reading, and pupils read the passages in the order 1 to 6. If a pupil finds any passage hard then its corresponding module booklet will be the starting lesson. For example, if a pupil does not confidently read aloud to the teacher passage 3 then its corresponding module 6 (as shown in the FS grid of assessment) will be the starting module. This will regroup pupils and the new grouping will have pupils with same modules or at the same reading age according to the modules.

There are 33 modules to be completed and if a pupil's starting module is 6 then in the complete programme the target modules will be 6 to 33. The modules are graded in 6 sets and each set has a pack of 5 booklets with different titles (11 A4 size pages per book). After completing each set pupils are assessed individually on the entry test to see their readiness to start the next module set. The

modules are also graded according to the NC Levels for reading at 1b to 4b Level. FS claims that after module 21 pupils will be unlikely to need FS intervention and they can join the mainstream classroom without any difficulty.

The contents of each module focus on the review and practice of Speed Sounds. The teacher should practice Speed Sounds through sound charts and sound cards before every reading activity. In a module each booklet has nine reading activities based on the text (fiction or non-fiction). The topics of the stories are according to the interest level of pupils in Year 7. For each story there are Green Words (decodable words, e.g. thank, munch, great), Red Words (non-decodable words, tricky words, e.g. the, could, were) and Challenge Words (new words). All of these words are read as Sound Talk and discussed by the teacher with pupils. Each story is read three times in one FS session in order to maximise the talk about text and support fluency and decoding. Before the story is read the teacher explains the new vocabulary. Each story has an introduction to be read aloud to the pupils and to develop pupils' interest in the story through discussion. The teacher first reads the story aloud and then pupils read it in pairs. There are questions at the end of each story; pupils read the questions and find answers in the text. Pupils are asked to read the story a third time with expression and a storyteller voice. The teacher points at the punctuation in the text. This happens in every FS session except for the sessions for assessment and entry test. The teaching is very text-centred and the teacher leads every activity of the session.

There are writing activities in each module which include spelling checks, mime punctuation, building sentences for composition, editing your own work and composition writing. Pupils follow the teacher and do the partner work, independent aloud reading and writing only on teacher's instructions.

### Teaching strategies and classroom management style

FS has its specific set of teaching practices for classroom management and pupil participation. The practices are introduced as rules of FS session and are embedded in general communication between teacher and pupils. The developers describe the process as follows:

- When teacher speaks (point towards teacher herself as 'my turn' (MT) and then sound out)
- Pupil repeats (point to the pupils saying 'your turn' (YT) where pupil repeats what the teacher sounds out)
- Turn to your partner (teacher wants pupil to practise with their partners: 'Talk to your partner' (TTYP))
- Stop signal (the teacher raises their hand to stop all the activity and sometimes counts down from 5 to 1)

In every class these routines are followed in order to manage communication in small-group learning sessions. Pupils become so used to this teaching style and communication that after a few sessions the teacher's silent signals are enough to pass the instruction for MT, YT and TTYP.

As described above FS is a structured intervention with clear steps for implementation and a process of regular assessment. The complete FS package includes module sets, assessment charts, magnetic sound cards, Speed Sound cards, sound charts and poster, lesson plans, phoneme pronunciation guide DVD for teacher, teacher training books and handbooks which provide detailed support on implementing FS. There are also two-day staff training workshops offered by Read Write Inc. where experienced FS trainers provide the hands-on training and teacher support visits in the actual classrooms.

As observed during the FS training and FS classroom sessions, teachers themselves need to have a strong grip on their phonics awareness and should have a good practice of teaching phonics if they are using FS. The teaching and classroom management strategies of FS are inspired from reception years and primary school teaching. The FS teacher is expected to bring in body language, praise

words, fun discussion and drama that can fix pupils' attention on the teacher. For a secondary school teacher, adopting the FS teaching style can be challenging as it demands a lot of dramatic performance to keep children's attention focused on the teacher.

In the process evaluation, it was observed that the schools that had been using FS already were more ready to implement the sessions as the teaching staff was used to FS methods. However, the schools new to the FS intervention adopted the FS teaching style gradually.

Control group activity: Pupils in the control group continued the usual classroom activities. The FS materials were not made available to the pupils in the control group or in usual classes.

## Sample size

This is a school-run trial in which the target sample was determined by individual school leads along with consultation with their cluster heads. Initially the schools proposed a rough estimate of 500 pupils all across 10 schools, which would traditionally be considered sufficient to detect a minimum effect size of around 0.26 (Lehr's approximation). This was dependent on English KS2 results for the fresh intake of pupils by these schools. The evaluators gave regular advice on pupil recruitment, selection criteria and randomisation. The total sample eventually included 433 pupils from three clusters with 10 secondary schools. The selection was based on English KS2 scores and pupils selected in the target group were at Level 4c and below in English.

## Randomisation

The evaluators conducted a tailored workshop on conducting an RCT and the process of randomisation before the trial began in September 2013. The meeting was attended by the 3 cluster leads and 10 school leads. In the workshop, methods of randomisation were suggested by the principal investigator and discussed with the cluster leads. The emphasis was on keeping the process fair and unbiased by completing the randomisation in one single phase.

The schools sent their identified target group and the background characteristics of pupils to the evaluators before they randomised the pupils into treatment and control groups. In each participant school the target groups were identified on the basis of KS2 assessment data on performance in reading. The schools used simple methods such as taking the names from a pot which contained chits for the pupils, a computer pseudo-random number generator, or a shuffled pack of card as demonstrated in the workshop. One of the schools initially identified a target group on the basis of KS2 scores, and TA assessment of reading. The pre-test was conducted on a target group of 18 pupils but the school leader was not satisfied with the target group and after consulting with evaluators they re-identified 18 pupils and conducted the pre-test again. In fact, the new target group had only two pupils replaced with the new ones. In one of the schools four more pupils were identified in the target group once the pre-test was conducted on all pupils. The new four were pre-tested and the evaluators were told that these pupils had been randomised and allocated to groups. In another school eight pupils were added to the target after the date of pre-test. The evaluators were informed that they were allocated to the groups after the usual process of randomisation.

Despite confirmation of a fair process of randomisation, it was observed by the evaluators that pupils with the lowest reading age or with special educational needs were more likely to be in the treatment group than in the control group. This process of extended randomisation may have created opportunities for selection bias. The waiting-list design was partly intended to prevent any subversion. Although conducted by the school leads, the evaluators were monitoring and strongly advising on the process throughout the period of intervention. The evaluators wanted to ensure independence of allocation by enforcing steps such as receiving the names of target pupils and matching the names after group allocation.

## Analysis

The analysis used here was based on 'intention to treat', meaning that all pupils originally identified as eligible were tested and their outcomes analysed, regardless of the time actually spent on the intervention. The impact of the trial is represented by the effect size (Hedges'  $g$ ) for the gain scores from pre- to post-test using the New Group Reading Test (versions A and B). Further analyses include using the same 'effect' size calculation for sub-groups such as FSM pupils, EAL pupils, boys and those with identified special needs.

Two simple multivariate regression models were also created. The two outcomes were the gain scores, and the post-test-only results. The possible predictors added to the models included the pre-test score, the pupil age in months at pre-test, the precise time period between pre- and post-test for that individual, and the known pupil characteristics such as sex, FSM eligibility, ethnicity, SEN and EAL. These predictors were all added to the model in the first step. Once these values had been accounted for, the treatment group (a binary variable) and the number of sessions attended (dosage) were added as a second step.

Readers may wish to note that significance tests and confidence intervals are not presented in this report. These do not work as intended (Carver 1978), are almost always misinterpreted (Watts 1991), and can lead to serious mistakes (Falk and Greenbaum 1995). Above all, they take no account of sample quality or attrition (Lipsey et al. 2012), being predicated on complete random samples of a kind never encountered in real-life research (Berk and Freedman 2001). This kind of explanation should no longer be necessary; rather those who still use such approaches must explain what their cited probabilities could possibly be probabilities of.

## Process evaluation methods

The process evaluation was conducted in cooperation with the cluster leads and the school heads who were in charge of the intervention programme in their respective schools and at the same time leading the conduct of the trial for evaluation purposes. The evaluation team undertook the process evaluation to ensure that the school leaders were following the protocol of the RCT. The information on the process reported here is twofold, relating to both the process of the intervention as observed in the actual school settings, and the management of the trial by the cluster leads and school heads.

Two workshops were conducted with the school leaders in which the protocol for conducting the RCT was discussed. The first meeting was in the workshop format in which the lead evaluator presented the design of RCTs, and discussed the required preparation for conducting the intervention along with the trial. The second meeting was conducted before the post-test phase in which the evaluators explained the conduct of the test process and discussed with school leaders how to calculate and interpret the results of the tests.

The project leaders arranged FS training for their staff members, implemented and monitored the intervention, and collected the formal records and the views of pupils. The evaluation team members observed the training in one of the three clusters. Implementation of the intervention in progress and testing after 22 weeks of the intervention was also observed, in all clusters. The intervention in progress was observed at two time intervals: one in the beginning when the intervention started and the second when the intervention was nearing its completion phase. Altogether evaluators conducted 10 FS session observation visits and 3 visits for test observations in the schools.

Read Write Inc. website materials on FS and documents relating to the intervention were read to understand the aims and methods of FS intervention. In the schools, evaluation team members conducted 14 face-to-face interviews with staff, 15 individual and many group discussions with pupils, and 7 interviews with project leaders. These interviews were usually conducted without a formal schedule, and arose as the situation allowed. Interviews generally took place with FS teachers and



school leaders both before and after observed sessions. The observations of staff training and implementation of the programme in action were as simple and integrated and non-intrusive as possible. The schedule of visits was agreed with the school leaders and some interviews were arranged at that time. The interviews and field notes were part-transcribed and shared between the evaluation team. Transcription was based on substantive content alone, since they were not to be used for anything like conversational analysis. None of the schools staff members' and pupils' names have been used or will be reported.

The process evaluation was useful in assessing fidelity to treatment. The perceptions of school leaders, staff members and pupils provided indications of any resentment or resistance to the programme. They are also useful in identifying potential issues or barriers which could be addressed for any future scaling up, as reported below

## Impact evaluation results

### Timeline

#### July – August 2013

Schools initially identified pupils for the target groups as soon as they received the KS2 data from the primary schools. A meeting was also held in which the evaluators presented a workshop for all interested school staff on conducting an RCT. During this time school leads also prepared for the intervention by adding resources to the libraries. The list of pupils in the target group was also shared with the evaluators. The school leaders also sent the background characteristics of the pupils.

#### September 2013

The NGRTA pre-test was conducted, followed by the randomisation of the target group. Opt-out consent from parents was also sought by the school leaders in this month. The schools started the intervention in this month right after the pre-test were conducted. However, one school asked to re-randomise due to reported errors in identifying eligible pupils.

#### October 2013 – January 2014

The intervention continued except during the Christmas break in December 2013. The evaluators conducted visits in the participating schools for the process evaluation.

#### February 2014

A second meeting was conducted with the school leaders on how to analyse and report their own results. The NGRTB post-tests were conducted in the last week of February 2014. The test completion rate was nearly 65% in the first week of the specified time for the test window. By the second week, 95% of the pupils had completed the test.

#### March – May 2014

Some schools completed the testing in the first week of March 2014. However, there were some pupils who had moved and needed following to try and complete the test in other schools. The new schools were approached with the help of school leaders. The context of the project was explained to the schools and it was requested to conduct the NGRTB for each pupil. The communication with the new schools was all done via email and telephone. Generally these schools were found very cooperative in conducting the stand-alone test, or explaining why it was not possible.

#### June 2014

Pupil background data were updated and outcome measures were analysed by schools and then by the independent evaluators. The process evaluation was written up.

## Participants

### School characteristics

Three cluster heads, in different regions, independently proposed conducting the FS intervention and an evaluation to the EEF. There was no developer involved in any of these three proposals and the number of schools involved in each cluster was substantial enough to conduct the evaluation. It was decided that school clusters should run the trial with advice from evaluators, and then aggregate the results for the final analysis. The 10 schools involved were located at three different locations in Harlow, Holderness and Telford (Table 1).

**Table 1: The number of schools and pupils in the trial**

Groups	Total number
Schools recruited	10
Schools participated	10
Pupils in the target group	433
Pupils included in the final analysis	423

All schools were urban, mixed secondary schools, with some diversity (Table 2). The proportion of FSM-eligible pupils was over 30% in four schools. A selection of 433 eligible pupils was made by scrutinising nearly 3,000 pupils enrolled in the 10 secondary schools. In the final analysis 10 pupils dropped out and could not be followed for the post-test due to reasons such as: left country, home schooled, from travellers group, did not attend the school after the first initial days and did not provide details of the destination school. According to the drop-out figures in the groups, 6 pupils were in the treatment group and 4 pupils were in the control group. The attrition is less than 3 per cent of the total sample selected for the study.

**Table 2: School characteristics**

Category	Phase	Age range	No. of Pupils	% SEN	% FSM	% EAL	Average KS2 point score 2013	OFSTED Effectiveness 2012–2013
Academy converter	Secondary	1116	1125	8.9	34.3	9.6	26.8	Outstanding
Academy sponsor-led	Secondary	1116	1200	NA	NA	NA	NA	Special measures
Academy-converter main-stream	Secondary	1116	985	15.2	26.7	2.8	26.4	Good
Academy-converter main-stream	Secondary	11-18	978	8.1	6.2	22.7	28.1	Inadequate
Academy-converter main-stream	Secondary	11-16	1022	20.8	33.9	9.4	25.9	Good
Foundation School	Comprehensive	11-16	606	11.4	34.3	15.7	27.3	Satisfactory
Academy-converter main-stream	Comprehensive	11-18	481	27	56.5	5	26.5	Satisfactory
Community school	Comprehensive	11-18	866	7	30.2	0.5	26.9	Requires improvement

<b>Community school</b>	Comprehensive	11-18	1766	7.7	17.5	1.3	28.1	Requires improvement
<b>Community school</b>	Comprehensive	11-18	1155	4.9	22.9	0.4	28.2	Good

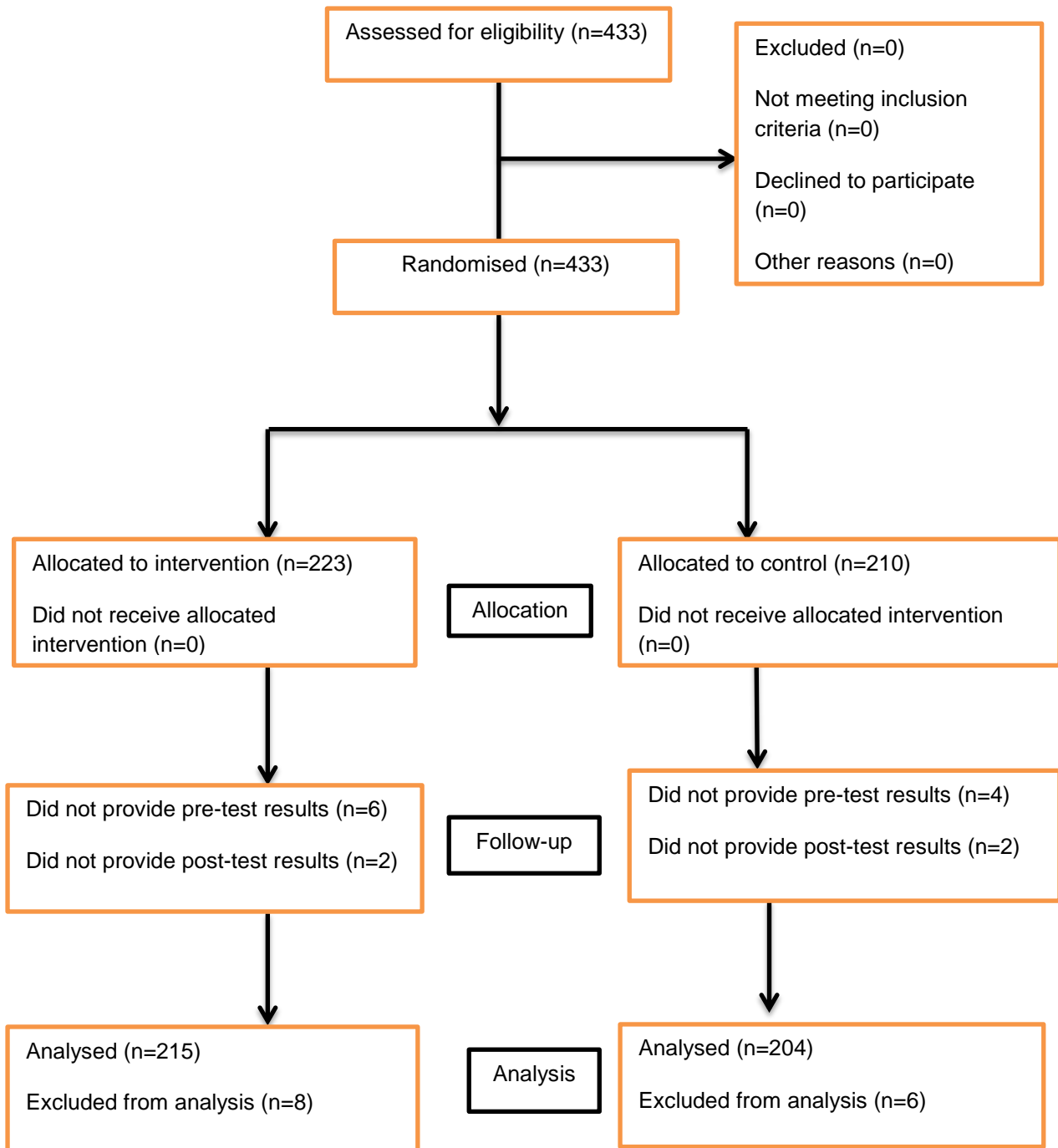
### Pupil characteristics

The pupil characteristics for the achieved sample and for each group are provided below (Table 3). In order to keep the pupil selection and group allocation procedures unbiased the following characteristics were considered only after the randomisation procedure. The results are reasonably well balanced in terms of background characteristics other than ethnicity.

**Table 3: Percentage of participants with specific characteristics in each treatment group**

	FS %	Control %
<b>Male</b>	54	50
<b>Female</b>	46	50
<b>FSM</b>	34	38
<b>SEN</b>	25	22
<b>EAL</b>	4	4
<b>Non-White UK</b>	15	8

**Figure 1: Participant response flowchart**



## Outcomes and analysis

The overall analysis could not include 14 pupils whose pre- or post-test data was missing, because they were ill or suspended at pre-test, and had moved schools and were either not contactable (having left the country, for example) or their new schools were not willing to deliver the post-test. The evaluation team along with the school leads tried to approach all the missing pupils to complete the post-test. There were eight pupils who completed the test as a result of this follow-up. The pupils who were dropped were those who could not be followed due to having left country, being home schooled or never having attended a participating secondary school after the initial few days. The missing pupils constitute less than 3% of the total sample randomised. They cover a range of pre- and post-test scores which are unremarkable in relation to the overall mean scores for all pupils taking these tests (except for one). There is no reason to suggest that this low level of attrition was biased in terms of test scores or allocation to group (Table 4).

**Table 4: Missing pupil scores and their allocated groups**

Group	Pre-test	Post-test
Control	-	232
Control	-	284
Control	224	-
Control	228	-
Control	252	-
Control	284	-
Treatment	-	300
Treatment	-	304
Treatment	4	-
Treatment	212	-
Treatment	248	-
Treatment	308	-
Treatment	312	-
Treatment	320	-
<b>Overall mean scores for trial</b>	263	285

Having been randomised by the schools, the two randomised groups were found to be not balanced in terms of the pre-test at the outset (Table 5). The pre-test scores of the intervention group are considerably lower on average, which may be evidence of interfering in the randomisation process by the schools. This is also suggested by the larger number of cases randomised to the treatment group. Therefore, gain scores are appropriate for most of the remainder of the analysis. The gain scores are the differences for each pupil between their post-test raw score and their pre-test raw score, and so should be largely independent of this initial imbalance. Using gain scores, FS had a small positive impact on reading comprehension. However, the initial imbalance means that this result must be more tentative than if the randomisation had been more effective. The difference in outcomes could be due to regression to the mean, for example, since the pupils with the lowest scores have the most opportunity for rapid improvement. The process evaluation pointed out several issues that could have led to the initial imbalance in the grouping.

**Table 5: Overall reading gain scores**

	N	NGRT-A	Standard deviation	NGRT-B	Standard deviation	Gain score	Standard deviation	'Effect' size
<b>FS</b>	215	251.8	65.4	279.5	59.9	27.5	47.7	+0.24
<b>Control</b>	204	274.2	58.2	290.6	53.3	16.7	42.1	-
<b>Total</b>	419	262.7	62.9	284.9	57.0	22.2	45.3	-

(Note: very similar results were obtained using the standardised age scores in place of the raw overall reading scores)

While the result in Table 5 could be linked to the initial imbalance between the groups, it is relatively unlikely to be due to either chance or bias caused by the dropout of 14 cases. For example, taking the mean gain score for the treatment group and adding one standard deviation creates a score that would be counterfactual to the scores in the control group. It would take over 35 such scores added to the existing control group to eliminate the effect size reported in Table 5.

Table 6 shows even more clearly why gain scores are needed. Because the treatment group had worse test scores at the outset, a post-test-only analysis would suggest that the treatment was damaging even though the FS group made more progress. The post-test-only scores would have been the final result if this trial had not chosen to conduct the pre-test on all of the pupils.

**Table 6: Overall reading post-test scores**

	N	NGRT-B	Standard deviation	'Effect' size
<b>FS</b>	215	279.5	59.9	-0.19
<b>Control</b>	204	290.6	53.3	-
<b>Total</b>	419	284.9	57.0	-

In order to help assess whether the higher gain scores for the group with the lower initial score could be due to regression to the mean, the analysis was repeated with only low-scoring pupils (Table 7). The result is based only on those pupils with a pre-test score below 258 (the overall mean). There are, as expected from Table 6, more low-scoring pupils in the treatment group than in the control. They have a slightly higher average gain score than the low-scoring pupils in the control group, but the 'effect' size is only +0.04. This means that if the intervention had any substantial impact then it was with the higher-attaining pupils initially. It also suggests that the result is not principally due to regression towards the mean.

**Table 7: Overall reading gain scores for low-scoring pupils**

	N	NGRT-A	Standard deviation	Gain score	Standard deviation	'Effect' size
<b>FS</b>	140	254.2	52.6	37.9	49.1	+0.04
<b>Control</b>	93	259.7	54.8	35.7	45.3	-
<b>Total</b>	233	256.4	53.4	37.0	47.5	-

The initial imbalance must make the rest of the results slightly more tentative. However, it is still useful to see the impact of FS at the sub-group levels in the complete sample. Table 8 suggests that FS is effective for FSM-eligible pupils.

**Table 8: Overall reading gain scores, FSM-eligible pupils only**

	N	NGRT-A	Standard deviation	NGRT-B	Standard deviation	Gain score	Standard deviation	'Effect' size
<b>FS</b>	52	243.1	54.2	271.6	59.8	28.5	49.2	+0.24
<b>Control</b>	52	264.7	58.4	283.2	52.5	19.5	38.5	-
<b>Total</b>	104	253.9	57.1	277.5	56.2	24.0	44.2	-

Tables 9 and 10 present the results of two further sub-groups of the sample. The results show that FS is associated with gains for both girls and boys. However, girls have shown more improvement than boys.

**Table 9: Overall reading gain scores, boys only**

	N	NGRT-A	Standard deviation	NGRT-B	Standard deviation	Gain score	Standard deviation	'Effect' size
<b>FS</b>	140	243.1	69.2	269.6	64.3	26.0	49.0	+0.17
<b>Control</b>	121	265.7	64.6	282.3	56.5	17.0	41.3	-
<b>Total</b>	261	253.6	67.9	275.5	61.0	21.8	45.8	-

**Table 10: Overall reading gain scores, girls only**

	N	NGRT-A	Standard deviation	NGRT-B	Standard deviation	Gain score	Standard deviation	'Effect' size
<b>FS</b>	75	268.0	54.3	298.2	45.2	30.2	45.4	+0.31
<b>Control</b>	83	286.6	45.0	303.0	45.7	16.4	43.5	-
<b>Total</b>	158	277.8	50.3	300.7	45.4	23.0	44.8	-

Table 11 presents the R values for the two regression models, each based on two steps. In step 1, the pupil background and pre-test scores are included, and then in step 2 the binary variable for being in the treatment group or control is added. The model is better at explaining variation in the post-test outcome than the gain score outcome. But for both models the bulk of the variation that is 'explained' by the variables in the model is explained at step 1. Once pupil background and prior attainment is accounted for, very little difference is made by knowing whether a pupil was in the treatment group or not. This model is not, in itself, any test of causation but it does provide a caution about the strength and importance of the intervention in relation to pupil characteristics. As the groups were unbalanced in terms of initial scores, this alone could explain why the R value for the treatment group is low.

**Table 11: Variation explained (R) in two-stage regression model, using two possible outcomes**

	Gain score outcome	Post-test outcome
<b>Step 1 – background and prior attainment</b>	0.54	0.74
<b>Step 2 – intervention and dosage</b>	0.54	0.74

For completeness, Table 12 presents the coefficients for all explanatory variables retained in either model. The largest of these by some way are the precise age in months when taking the test, and the pre-test score.



**Table 12: Standardised coefficients for the regression model in Table 11**

	Gain score outcome	Post-test outcome
<b>FSM</b>	+0.01	+0.00
<b>Sex (female)</b>	+0.08	+0.07
<b>SEN</b>	+0.06	+0.05
<b>EAL</b>	+0.01	+0.01
<b>Ethnicity (White UK)</b>	-0.06	-0.05
<b>Age at pre-test</b>	-1.04	-0.82
<b>Age at post-test</b>	+0.98	+0.78
<b>NGRT A (pre-test)</b>	-0.56	+0.66
<b>Step 2: Treatment (or not)</b>	-0.03	-0.03

*(Note: the precise period between pre- and post-test was not related to either outcome)*

## Cost

The direct cost of FS includes staff training and the materials for teaching. The cost of two-day staff training was £2,750 for 45 staff members in all schools. In order to have FS-experienced trainers visit classrooms and provide supporting feedback to the teachers the cost is £3,000 for six days. These days can be spread across the whole term. The teacher's kit includes all of the materials required for teaching and the cost per kit is £138. Children need module sets for reading and writing activities. The cost for one complete module set for one child is £95.90. The resource material for FS is published by Oxford University Press. Module booklets need to be purchased on a regular basis.

The cost of staffing to assist teachers depends largely on each school's initial position. So, for example, a school with four relevant teachers to be trained and 20 pupils per year group, would have to pay £2,316, or around £116 per pupil.

The schools have to invest a reasonably large sum of money in this intervention. The schools in this project have used a major part of their EEF funding in purchasing FS resources. The cost for extra teaching assistant staff was also covered in the project budget for individual schools. The heads reported that without this sum of money from EEF the schools could not possibly afford the cost of this intervention.

## Process evaluation results

### Implementation

#### **What are the necessary conditions for success of the intervention?**

FS is a teacher-led structured intervention that targets reading improvement through phonics awareness and regular practice. The main aspect of FS is teaching pupils in small groups of homogeneous reading ability. The groups are created on the basis of FS assessment material and the initial screening determines the starting point for each group. The intervention has specialist resource material which covers a systematic plan of learning decoding followed by reading comprehension. The FS-specific teaching style is a core element of this intervention which encompasses teacher's passion, praise for pupils and a dynamic pace for the lessons. The classroom management and teacher-pupil communication techniques are prescribed in the training and teacher's handbook. This suggests that this intervention cannot be successfully implemented on its own, without FS training or FS teacher's handbook.

FS training in the schools was given by experienced and professional trainers. In the two-day training, the concept of FS teaching through phonics was discussed along with FS teaching style and the rules of classroom management. The training also provided knowledge about phonemic awareness, how teachers should use modules for practising decoding, and pupils' assessment and grouping in the classroom. Sixty-five school teachers and teaching assistants attended the FS training, the majority of whom had no previous experience of using FS.

#### **Barriers to delivery**

FS is widely used and is reported to be an easily adaptable intervention that fits in with the school schedule. However, the concept of FS essentially depends on the teacher's experience and confidence in teaching through phonics. For secondary school staff, this seems highly dependent on FS teacher training and feedback from experienced FS teachers or trainers. FS demands teacher engagement with complete control of the classroom just like in a primary school setting. The format and teaching style of the primary school level seems to be a slight challenge to adopt for the secondary school teacher who would generally have a different teaching style. One head reported that one very experienced secondary school teacher and head of department refused to adopt this teaching style and signed off from the training. This might be a factor again in any larger-scale rollout of the intervention.

One school leader reported that parents initially questioned teachers regarding their child's involvement in the FS intervention saying that it was very basic and at primary school level. The school leader organised a meeting and explained the purpose of the intervention, which was then understood by the parents. It was reported that some pupils on the intervention also showed resistance to do basic phonics practice as they felt patronised. The teachers reported that they managed to overcome pupil's resistance through private discussion with them.

The pupil grouping was reported as a difficult decision when teachers had different sets of scores for pupils' performance. At the time when teachers grouped pupils they had pupils' KS2 scores in English, pre-test NGRT-A and the FS screening test scores but only used the FS screening tests. The decision of grouping children could be entirely different if the teachers had used NC scores or the pre-test NGRT-A scores because in each case there were some disparities in pupils' reading age. However, the protocol of FS intervention required grouping according to FS screening tests and pupils were therefore grouped accordingly. The pupil grouping was followed by the pre-test and randomisation process.

**Is the intervention attractive to stakeholders?**

Pupils on FS were observed during the sessions and some of them were interviewed after the sessions. The dynamic FS teaching style kept pupils engaged throughout the sessions. Pupils seemed to have enjoyed the practice of phonics through various mnemonics. The sessions included one FS teacher and one FS-trained TA so the pupils were seen to be given a lot of support and individual attention which they might not have received in the whole group. Many pupils reported that they preferred coming for FS sessions rather than going to other lessons

As above, schools generally received positive feedback from parents when they were informed about participation in the FS intervention. However, there were parents who raised concerns about their children's participation in such a basic programme. A meeting with teachers and parents helped them understand.

FS brings teaching resources and useful lesson plans in the classroom. The FS training includes a lot of information for teachers regarding phonics teaching programmes which has been reported to be useful also for teaching literacy in general. The school leaders reported that the intervention targets pupils who were at risk of falling behind so having FS-trained staff members in the school could immediately support these pupils according to their needs.

## Outcomes

### Perceived outcomes

One of the outcomes commonly reported by teachers was that FS provided positive results for pupils who have learning difficulties. Pupils' confidence in reading and class participation increased in and following FS sessions. The teachers considered that small-group teaching helped many pupils to develop their confidence and a better relationship with teachers, which may be unrelated to the specific nature of the intervention (several EEF evaluations of small-group literacy intervention have found similarly small positive effect sizes). The teachers often reported that pupils who were quiet or disruptive in whole class groups were more focused and confident in FS sessions.

Teachers found the implementation of FS very prescriptive. Some teachers reported that this structured plan and teaching style were crucial to the overall impact of FS. However, other teachers were not very keen on prescription at every point of the FS sessions.

Pupils in the intervention generally liked FS sessions. However, some pupils felt that what they learnt in FS sessions had been taught in their primary school and they did not need to repeat such basic things. Sometimes the reading tasks were very easy for them. This feedback from some pupils is based only on their perception of learning and their teachers did not usually agree with their views.

The FS attendance records showed that pupils were attending the sessions very regularly. Although many of them had to leave regular classes, the FS sessions were well attended. Some pupils reported that they preferred coming for FS sessions rather than attending regular classes because they found other lessons boring.

The results of NGRT were based on pupils' performance on the online NGRT which adapts the level of challenge according to pupils' initial responses. Pupils did not practise doing the screen-test. According to the school leaders the NGRT was a challenging test of reading comprehension for pupils. The evaluators observed during the test process that the following main sections in reading comprehension were assessed in NGRT:

- Context comprehension (inference and deduction)
- Writer's purpose and viewpoint
- Writer's use of language
- Organisation of texts
- Social, cultural, historical and traditional aspects in the given piece of text

The format of the test was multiple-choice questions. It was quite possible that some pupils were just guessing the right answers. However, in terms of appropriateness of the reading test content and the level of challenge for pupils in secondary (Year 7), NGRT fulfils the criteria of an appropriate assessment tool.

### Fidelity

FS is a 33-week daily one-hour intervention but for this trial the total duration was reduced to 22 weeks for three days a week. This adaptation was agreed with all of the cluster heads and school leaders so that all schools in this aggregated trial implemented FS at standard levels of dosage and with the least disruption to other activities in the schools.

The process of FS as demonstrated to teachers in the training was generally observed as being followed in the school visited. The staff covered the main aspects of the intervention successfully. The following core features of FS were commonly observed in the participant schools:

- Screening test for pupil grouping
- Phonics awareness and practice through mnemonics
- Modules entry test and reading and writing practice
- Pupil's participation
- Classroom management rules

The process of FS is very prescriptive in terms of teacher behaviour, language and communication with pupils. In the sessions it was observed that the teachers tried to follow all the rules of the intervention. However there were still variations, as teachers have their own style and personality. FS trainers were more experienced and they could almost mimic the developer's style of teaching and classroom engagement. This level of replicating a style is perhaps not possible for every teacher. The ones observed were certainly enthusiastic and caring.

The NGRT post-test was pre-arranged and during the last week of the trial all schools conducted the post-test. The post-tests were conducted with batches of pupils as the test was online and required to be completed on a computer connected to the Internet. The schools arranged groups of pupils to take the test and the evaluators ensured that pupils in the treatment and control groups were given equal time, space and test instructions.

## Formative findings

### Are there any ways that the intervention can be improved?

The limited suggestions for improvement of the FS intervention are based on what evaluators observed, and what school leaders, teaching staff and pupils experienced and reported.

### Teacher development

FS provides comprehensive teacher training on teaching English through phonics. This training is a very useful capacity-building programme for the secondary school teacher who may be a literature specialist and not used to such a basic approach. However, the concept of teaching proposed in the training can still seem alien to those that the teacher uses successfully in mainstream teaching. Adopting the intervention means that the FS teachers must not use their own techniques of spellings, word recognition and comprehension. The letters of the alphabet must not be read by their names, only their sounds, for example. FS could be an effective method of teaching in itself. However, this intervention is only a catch-up programme and once completed the pupils are supposed to join the mainstream groups. Teachers in the mainstream classroom would not necessarily apply these rules but stick to their own teaching methods and teaching style. The FS catch-up programme needs to consider that secondary school mainstream teaching and classroom management styles are very different from those that FS proposes. They could extend the advice and practice for teachers to prepare pupils to merge in the mainstream secondary school settings.

### Perceived effectiveness for SEN pupils

It is widely assumed that the FS programme is effective for pupils who have learning difficulties, but the evidence base is weak (i.e. it is untested). The developers could address the aims of FS if their target is pupils who have learning difficulties. If it is for all pupils at risk then future evaluations must include a variety of pupils who need catch-up interventions.

### **Can school leaders conduct RCTs?**

During the evaluation of this project, the evaluators observed the process with this question in mind: can school leaders conduct independent RCT or not? In terms of managing the intervention school leaders were free to make some decisions but overall their decisions were guided by a highly structured FS concept delivered during the training and support visits by the trainers.

School leaders had taken complete responsibility for the implementation of intervention in their schools. This could support the idea of independent RCT conducted by the school leaders without the involvement of intervention developers. However, the implementation without the developers' training and support staff visits would not have met the protocol of the intervention. Although supported with teacher's guide and handbook materials, the idea of FS cannot be independently implemented. After the training the support staff visits provided feedback in the real classroom sessions and demonstrated in front of teachers. This partial involvement of FS trainers could not be avoided and the schools therefore conducted the trial with the feedback of FS trainers on grouping and pupil selection.

Cluster heads and school leaders being in charge of the intervention and the trial management helped reduce attrition as they took ownership of the project with support from the evaluation team.

In terms of pupils' selection and randomised group allocation, school leaders were observed to be biased towards pupils with special education needs. FS was perceived to be an intervention that works for pupils with very low reading age and its effectiveness was thought to be seen only on pupils with the lowest reading levels. Pupils having special education needs are more often the ones with the lowest reading age. This understanding of FS may have led to unbalanced grouping of pupils in the treatment and controlled groups, with more SEN pupils in the treatment group.

The process of evaluation was guided and monitored by the expert evaluators. Communication between cluster leads and school leaders and the evaluation team was very regular and exchange of information was quite frequent. Two workshops were organised for school leaders so they could understand the process of research for evaluation. Feedback suggested that the meetings were useful in finding out school leaders' views of the research project and the challenges that teachers face in conducting small-scale research projects in their schools. The school leaders also stated that conducting the research project was an educational experience.

The independent evaluators aggregated the results of 10 schools, and conducted the final analysis of results and their reporting here. The cluster heads and school leaders were given guidance, materials and a template to develop individual school reports. In fact only one cluster head from Harlow completed a formal report and submitted the results of the trial for that cluster. Such analysis and reporting may be a widespread problem for schools attempting their own trials.

### **Control group activity**

Pupils in the control group continued normal lessons in English. In some schools FS-trained teachers were teaching these normal English lessons and therefore there is a small chance of diffusion to pupils in the control group. However, FS is a radically different approach only applicable for some (see above), and FS materials and modules were not accessible to pupils in the control group. In general, pupils in the control group continued usual activities in the school which sometimes involved other reading interventions and catch-up literacy programmes.

## Conclusions and implications

### Limitations

The design adopted for this trial does not establish the long-term impact of this intervention. Once the tests were completed there was no opportunity of follow-up to see if FS pupils were successfully merged with the mainstream group.

The schools participating in the trial had applied individually for funding from EEF to conduct the intervention. The schools are not a representative sample of a larger population in the area. The schools conducted the intervention regularly and followed the protocol of the intervention. There was no example of a school in which FS was partially implemented.

Since the FS developers were not directly involved in the application for funding there was initially no conflict of interest which could influence the results. However, there is a possibility that schools' enthusiasm to take part in a funded research project resulted in a kind of Hawthorne effect. The schools volunteered to conduct the intervention for the purpose of evaluation but all of them harboured a preconception that FS is a very effective intervention and that it improves pupils' reading age.

Random allocation was done by school leaders. However, the FS trainers may have had an influence in at least some schools. There seems to be a selection bias in the group allocation despite regular meetings with evaluators and monitoring of the process. The groups were initially very imbalanced, which could have occurred by chance but is highly unlikely. Some schools had asked to re-randomise after contact with the developers during training.

The randomisation was for individual pupils within schools, and pupils in both arms of the trial continued the activities in the school except for FS sessions. Students in the treatment group were taken out of the regular classes for FS but even then there is a danger of some treatment diffusion to the pupils in the control group. The teachers conducting FS were sometimes teaching normal English lessons as well.

The results of NGRT are based on pupils' performance on the screen test. NGRT is a computerised test which adapts the challenge according to the pupil's initial responses. This could be a challenge for some pupils or they might consider it as a game.

### Interpretation

The main objective of this RCT was to assess the gains in reading comprehension after using FS as opposed to the test performance of pupils in the control group. No school dropped out of the evaluation, FS was regularly administered, pupils in the control group were not given access to FS material or organised FS sessions, and the pupil attrition rate was no more than 3%. Therefore, the overall findings have a medium level of evidence that FS was an effective reading intervention with the effect size of +0.24. This is approximately equivalent to three months' progress in the reading age of pupils over the course of a whole year. The estimated cost of using FS for 33 weeks is mainly the cost of module booklets, which is £95.90 per child, plus the cost of a trained teacher and TA time for 33 weeks.

FS is a structured and developed intervention. It is tailored for pupils at risk and those who need to catch up with a mainstream group. It offers extra support for low-achieving pupils. It is not a whole-school programme, therefore is not relevant to pupils who are at their appropriate reading age level or above.

Individual schools have demonstrated that they can learn about RCTs and conduct most of the steps themselves without continuous direct supervision. However, there are some aspects that still require oversight by independent evaluators. This is even more important in inferior designs (such as school-level allocation) or where there is likely to be any conflict of interest. It is entirely understandable that schools naturally tend to be more concerned with the progress of each specific cohort than with the process of building a secure evidence base for others. Part of the evaluators' role was to judge whether, given expert advice, schools were able to conduct RCTs. In this respect, the answer has to be that schools are probably not in the best position to randomise their own pupils. The majority of schools in this trial did not analyse and report results, suggesting that those running trials in their own settings will continue to need the help of independent evaluators.

### Future research and publications

Fresh Start is a well-structured intervention with developed plans for training, resource materials and assessment. It targets improvement in reading and writing at the same level. This RCT has shown that schools can implement FS as a catch-up intervention for pupils on transition from primary to secondary, and the results are promising for pupils who are at risk of falling behind in literacy. However, the evidence achieved does not provide solid confirmation of the effectiveness of this intervention, because the results were obtained from such initially imbalanced groups. It is possible that some schools subverted the randomisation process, perhaps on the advice of the developers. The question of the actual effectiveness of FS would repay further work. The following research questions have also emerged.

What is the 'active' ingredient of FS? There are four main elements of FS: pupils' screening and grouping, teaching phonics, practice decoding and comprehension on modules, and following the prescribed classroom management rules. What is the most effective element of this whole process? Is it teaching in homogeneous groups of pupils in terms of reading age? Does it depend on practising decoding on modules? Or is it the atmosphere of learning created by FS teacher? Does it require all of these?

Do pupils successfully rejoin their mainstream classes? This is one of the claims of FS – that after completing 33 modules a pupil would not need any further intervention and could be merged with the mainstream class to continue their literacy activities. This is mainly a question about sustaining the impact of FS.

In conclusion, it is worth repeating that the intervention has promise, and that it was implemented successfully. However, for conclusive results on the effectiveness of FS a largely and independently randomised controlled trial would need to be conducted.

The evaluators plan to use the results of this trial as part of a paper in a peer-reviewed journal.



## References

- Berk, R. and Freedman, D. (2001) Statistical assumptions as empirical commitments, <http://www.stat.berkeley.edu/~census/berk2.pdf>, accessed 030714
- Brooks, G. (2003) Sound Sense: the phonics element of the National Literacy Strategy. A report to the Department for Education and Skills, DfES website, 20/8/03: [http://www.standards.dfes.gov.uk/pdf/literacy/gbrooks\\_phonics.pdf](http://www.standards.dfes.gov.uk/pdf/literacy/gbrooks_phonics.pdf)
- Brooks, G. (2007) What works for pupils with literacy difficulties? The effectiveness of intervention schemes, London: DCSF Publications
- Brooks, G., Harman, J. and Harman, M. (2003) Catching Up at Key Stage 3: an evaluation of the Ruth Miskin [RML2] pilot project 2002/2003. A report to the Department for Education and Skills, Sheffield: University of Sheffield
- Carver, R. (1978) The case against statistical significance testing, *Harvard Educational Review*, 48, 378-399
- Department for Education (2013) Statistical first release: National Curriculum Assessments at Key Stage 2 in England, 2013 (revised). DfE: National Statistics  
[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/264987/SFR51\\_2013\\_KS2\\_Text.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/264987/SFR51_2013_KS2_Text.pdf)
- Ehri, L., Nunes, S., Stahl, S. and Willows, D. (2001) Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis, *Review of Educational Research*, 71, 3, 393-447
- Falk, R. and Greenbaum, C. (1995) Significance tests die hard: the amazing persistence of a probabilistic misconception, *Theory and Psychology*, 5, 75-98
- Gorard, S. (2013) *Research Design: Robust approaches for the social sciences*, London: SAGE
- House of Commons Education and Skills Committee (2005) *Teaching Children to Read: Eighth report of session 2004-05*, <http://www.publications.parliament.uk/pa/cm200405/cmselect/cmmeduski/121/121.pdf>
- Johnston, R. and Watson, J. (2004) Accelerating the development of reading, spelling and phonemic awareness skills in initial readers, *Reading and Writing: An Interdisciplinary Journal*, 17, 4, 327-57
- Johnston, R. and Watson, J. (2005). The effects of synthetic phonics teaching on reading and spelling attainment, Edinburgh: Scottish Executive Education Department, <http://www.scotland.gov.uk/Resource/Doc/36496/0023582.pdf>
- Lanes, D., Perkins, D., Whatmuff, T., Tarokh, H. and Vincent, R. (2005) A survey of Leicester City Schools using the RML1 and RML2 literacy programme, Leicester: Leicester City LEA. (mimeograph)
- Lipsey, M., Puzio, K., Yun, C., Hebert, M., Steinka-Fry, K., Cole, M., Roberts, M., Anthony, K. and Busick, M. (2012) Translating the statistical representation of the effects of education interventions into more readily interpretable forms, Washington DC: Institute of Education Sciences
- OFSTED (2010) *Reading by six: How the best schools do it*, Ofsted: UK
- Rose J. (2006) *Independent review of the Teaching of Early Reading*, London: DFES  
[http://www.literacytrust.org.uk/assets/0000/1175/Rose\\_Review.pdf](http://www.literacytrust.org.uk/assets/0000/1175/Rose_Review.pdf)

Torgerson, C. (2003) *Systematic Reviews*, London: Continuum Books

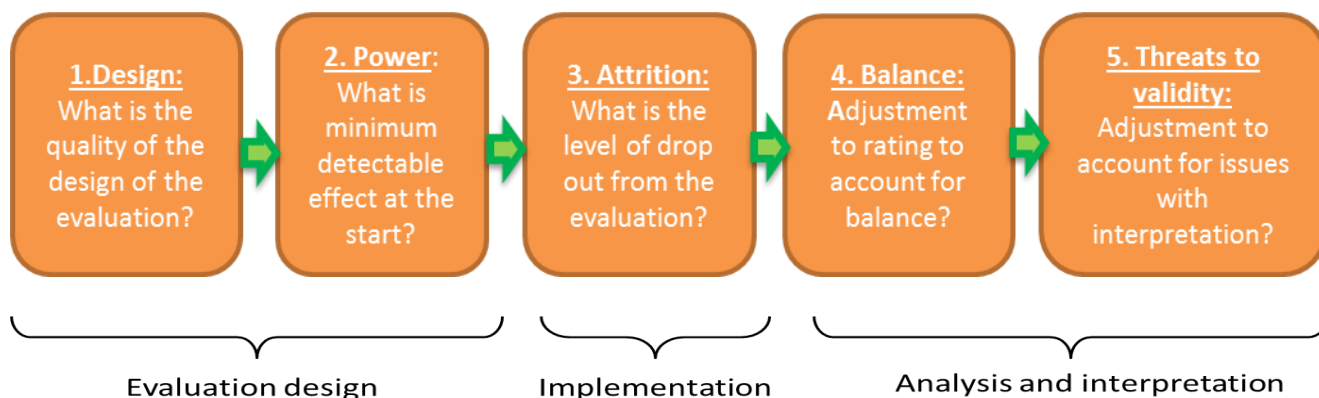
Torgerson, C., Brooks, G. and Hall, J. (2006) *A Systematic Review of the Research Literature on the Use of Phonics in the Teaching of Reading and Spelling*, London: DfES, Research Report

Watts, D. (1991) Why is introductory statistics difficult to learn?, *The American Statistician*, 45, 4, 290-291

What Works Clearing House (WWCH) (2010) *Sound partners*, US Department of Education: Institute of Education Sciences,  
[http://ies.ed.gov/ncee/wwc/pdf/intervention\\_reports/wwc\\_soundpartners\\_092110.pdf](http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_soundpartners_092110.pdf)

Wyse, D. and Goswami, U. (2008) Synthetic phonics and the teaching of reading, *British Educational Research Journal*, 34, 6, 691-710

## Appendix 1: Padlock rating



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

The final security rating for this trial is 3 . This means that the conclusions have moderate security.

The trial was designed as an efficacy trial, partly administered by schools, and could achieve a maximum of 5 . The study was relatively small and lost 2 padlocks. The attrition was extremely low. There was indication of substantial imbalance between the arms at baseline (ES of 0.24) that was potentially not due to chance. Therefore, the 2 padlocks were lost. There were a few threats to validity: the schools administered the randomisation without record; there was suggestion that the random allocation was subverted, and the teachers administered the testing. Therefore, the overall security was 3 .

## Appendix 2: Cost rating

Cost ratings are based on the approximate cost per pupil of implementing the intervention over one year. Cost ratings are awarded using the following criteria.

Cost	Description
£	<i>Very low:</i> less than £80 per pupil per year.
£ £	<i>Low:</i> up to about £170 per pupil per year.
£ £ £	<i>Moderate:</i> up to about £700 per pupil per year.
£ £ £ £	<i>High:</i> up to £1,200 per pupil per year.
£ £ £ £ £	<i>Very high:</i> over £1,200 per pupil per year.

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

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

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

This document is available for download at [www.educationendowmentfoundation.org.uk](http://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](http://www.educationendowmentfoundation.org.uk)