Education
Endowment
Foundation

Success for All

Evaluation report and executive summary
July 2017

Independent evaluators:

Dr Sarah Miller, Dr Andy Biggart, Dr Seaneen Sloan, Dr Liam O'Hare

Queen's University
Belfast

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

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

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

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

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

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

Danielle Mason
Head of Research and Publications
Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

p: 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 Queens University Belfast: Dr Sarah Miller, Dr Andy Biggart, Dr Seaneen Sloan, Dr Liam O'Hare.

The lead evaluator was Dr Sarah Miller.

Contact details:

Dr Sarah Miller

Address: Queens University Belfast, University Road, Belfast, BT7 1NN, Northern Ireland, UK
Email: s.j.miller@qub.ac.uk

Tel: 028 9097 5944

Centre for Evidence and Social Innovation at Queen's University Belfast
http://www.qub.ac.uk/research-centres/cesi/

Contents

Executive summary

The project

Success for All (SfA) is a whole-school approach to improving literacy in primary schools. Teachers receive training in areas including group learning strategies, phonics, and assessment, and are provided with structured daily lesson plans and teaching materials. There is also support for school leadership in areas such as data management, ability grouping, and parental engagement. In this trial, the programme was delivered over two academic years on a whole-school basis by teachers and classroom assistants who were trained and supported by the Success for All Foundation. In total, schools received 16 days of training and support. Although the intervention was delivered on a whole-school basis, the evaluation focused only on the outcomes of 1,767 pupils starting in Reception class, following them through until the end of Year 1. Fifty-four schools took part in this effectiveness trial. An implementation and process evaluation ran alongside the trial to explore implementation and delivery challenges. The evaluation took place between September 2013 and June 2016.

Key conclusions
1. Children who took part in Success for All (SfA) made 1 additional month's progress, on average, after two years compared to children in other schools. The 3 padlock security rating means that we are moderately confident that this difference was due to SfA.
2. Children eligible for free school meals (FSM) made 2 additional months' progress after two years, compared to FSM children in control schools. The smaller number of FSM pupils in the trial limits the security of this result, though combined with other findings in the report it provides some evidence that SfA does improve literacy ability for children eligible for free school meals.
3. Of the 27 schools receiving SfA, 7 dropped out of the programme completely, and 5 did not implement it to the minimum expected level. There is some indication that lack of engagement with the programme resulted in poorer implementation which may have affected the programme's impact.
4. Some schools found SfA prescriptive and adapted the content and delivery of the programme. However, there is some evidence that schools delivering SfA as prescribed by the developers saw larger gains in children's literacy after two years than those that completed fewer aspects of the programme.
5. Overall, schools that successfully delivered SfA were enthusiastic and valued the programme in their school.

EEF security rating

Security rating: 

These findings have moderate security. The security rating of the trial indicates how confident we can be that the additional progress experienced by the children in the trial was due to SfA and not any other factors.

This was an effectiveness trial, which tested whether the intervention could have an impact when delivered in a format that could be made available to a large number of schools. The trial was a well-designed two-armed randomised controlled trial, and the pupils in SfA schools were similar to those in the comparison schools. However, the trial has some limitations that reduce our confidence in the findings. 12% of children were not included in the analysis at the end of Reception, and 24% of children were not included in the analysis at the end of Year 1. The correlation between the pre-test and post-test was smaller than expected, which reduced the power of the trial to detect small effects.

Additional findings

The main analysis found no evidence to suggest that the intervention group had better literacy skills compared to the control group at mid-point (end of Reception class). However, at post-test (end of Year

1) the intervention group pupils had made an additional one month's progress on the Woodcock and the phonics check compared to the control group. Due to the missing data, a sensitivity analysis was conducted and this confirmed the result of the main (complete case) analysis.

An 'on treatment' analysis (excluding the data from the seven withdrawn intervention schools) suggested that there was some evidence of a small positive effect (one month's progress) in favour of the intervention group on the Woodcock literacy test at the end of Year 1. Schools that implemented the programme with greater fidelity saw larger impacts on the Woodcock literacy test and the phonics check at the end of Year 1, compared to schools with lower fidelity. These findings are exploratory, but together provide initial evidence that the programme is more effective when implemented as prescribed by the developers.

The subgroup analysis, although underpowered, suggested that the programme worked better for children who were eligible for FSM compared to their less deprived peers. Children with lower baseline attainment may also have received small benefits from the programme; however, this finding has limited security.

This is the third RCT of Success for All to be conducted and it is the first independent trial of the programme in England. The only other trials were based in the US and reported a positive effect of the programme, achieving effect sizes in the region of 0.15-0.30. The current trial has been unable to replicate these effects in an English context. Other less robust studies have reported mixed effects of the programme.

The findings from the process evaluation suggest that fidelity and quality of implementation varied between schools. The majority of teachers who took part in the process evaluation reported finding the programme prescriptive and some reported that they adapted the content and delivery. Similarly, SfA rated the quality of implementation of 15 intervention schools and only one reached the highest 'refined delivery' status after two years. In combination with the findings of the 'on treatment' and fidelity analyses these findings suggest that levels of fidelity may have had an impact on the overall effect of the intervention.

Cost

Over a three-year period the average cost per pupil per year is £61.90. This cost assumes whole-school delivery to a school of 275 pupils. SfA costs approximately £46,500 per school to implement with the whole school in the first year; costs reduce significantly in subsequent years. There is 2 days of training for all staff, and SfA provides 14 days of ongoing support over 2 years, tailored to the schools' needs.

Summary of main outcomes

Outcome	Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	P value	EEF cost rating
Woodcock at mid-point (end of Reception class)	Treatment vs. control	0.04 [-0.06, 0.14]	0 months	🔒🔒🔒🔓🔓	0.42	£ £ £ £ £
	Treatment FSM vs. control FSM	0.22 [0.01, 0.44]	3 months	n/a	0.03	£ £ £ £ £
Woodcock at post-test (end of Year 1)	Treatment vs. control	0.07 [-0.03, 0.18]	1 month	🔒🔒🔒🔓🔓	0.14	£ £ £ £ £
	Treatment FSM vs. control FSM	0.12 [-0.10, 0.34]	2 months	n/a	0.23	£ £ £ £ £

Introduction

Intervention

Success for All (SfA) is a multi-component training and support programme for primary schools that is underpinned by cooperative learning pedagogy. Developed in the US, it is a whole-school reform programme that includes the teaching of synthetic phonics, ongoing assessment, and ability grouping. In addition, there is a major focus on professional staff development, parent involvement, and school-wide structures which have been identified as important components of successful literacy programmes (Slavin et al., 2009). The theory of change that underpins SfA hypothesises that prevention and early, intensive intervention are the key to children's success both in school and in later life outcomes. The prevention element of the programme starts in Nursery and Reception class and draws on cooperative learning strategies, theme-based activities and story-telling to develop children's vocabulary and phonemic awareness. The early intervention aspect of the programme seeks to engage parents more fully in their child's learning and use tutoring strategies to address any emerging issues, before they become problematic. Cooperative learning strategies form a large part of the programme whereby children are put into mixed ability pairs or groups and teams are rewarded for each member's learning. Group membership is revisited and changed every term and from Year 1 onwards groups consist of children at the same reading level, regardless of what Year they are in.

Schools who implement the programme receive 16 days of in-school training and ongoing support directly from SfA-UK. Initial training is delivered by SfA-UK, a UK registered charity, and takes place on a whole-school basis (including all teachers and teaching assistants) over the course of two days at the beginning of the school year. The remaining 14 days takes the form of on-going in-school support. Day one of training focuses on cooperative learning, and on day two, staff are divided into groups depending upon the age group they teach. Schools receive support from SfA on an ongoing basis, tailored to the needs of the school. One member of experienced teaching staff in the school acts as the school's SfA coordinator. Their role is to work with all members of staff to help ensure optimal implementation. While the current trial focused only on outcomes for children in Reception class and Year 1, SfA was implemented on a whole-school basis in the intervention schools and so all teachers were provided with structured daily lesson plans and teaching materials (regardless of whether their class was involved in the evaluation or not). The trial recruited children who were starting Reception class and followed them for two years, to the end of Year 1, meaning that intervention children had a different teacher in Reception class and Year 1 and both teachers should have received the training and support from SfA.

Following training, the components of SfA are delivered by teachers and teaching assistants in the classroom, on a whole-class basis; however, the exact content of the training and materials, as well as the operational aspects of SfA, differ depending on stage of the curriculum. At the primary level, the programme is divided into three stages. Kinder Corner – aligned with the Early Years curriculum – is delivered in Reception class. Roots (for pupils working at the former National Curriculum level 2c) and Wings (for pupils working at the former National Curriculum level 2b) are both delivered with Years 1–6. The Kinder Corner programme is designed to be delivered over the course of the full school day. Roots and Wings involve 90-minute literacy sessions every day.

SfA was first introduced in the UK in 1997 and is targeted primarily at schools with high levels of deprivation. Some adaptations to the version of the programme developed in US were necessary for implementation in the UK, so that it aligned with the curriculum and to ensure appropriateness of language. These adaptations were made by a previous head of SfA-UK and are detailed in the programme manual.

Background evidence

The great emphasis that has been placed on the development of children's early literacy skills is very much related to its role as a gateway subject. Literacy skills are widely recognised as an important precursor to general academic achievement as well as in relation to broader participation in society. Longitudinal studies have also shown that children who fail to gain adequate basic literacy skills at an early stage are unlikely to catch up later (Brooks, 2007; Francis et al., 1996; Juel, 1988).

There has, however, been considerable debate and controversy over the best approaches for the teaching of literacy, especially among struggling beginning readers. This has been exemplified by what has been termed the 'literacy wars' between whole-language approaches and those that advocate the teaching of phonics. A systematic review limited to evidence from randomised trials concluded there was evidence that systematic phonics instruction was more effective than whole language or word approaches (Torgerson et al., 2003). Although the importance of the teaching of phonics has a strong evidence base, many reviews have concluded that it is insufficient on its own. Cowen (2003), for example, synthesized six major research studies which considered the early stages of learning to read and concluded that direct phonics teaching should not be taught on its own, in isolation from meaning and understanding. A number of national reviews, including the Rose Report, support this increasing consensus that a variety of approaches are required, including the use of systematic phonics, to support the literacy needs of all children (Rose, 2006; NICHD, 2000; Rowe, 2005).

The current impact evaluation focused upon providing an independent assessment of the effectiveness of the main SfA programme in improving struggling readers' literacy skills in English schools. The SfA programme has good evidence of effectiveness within the United States where a large number of evaluations have recorded positive improvements in children's literacy with pooled average effect sizes of around +0.5 (Slavin et al., 2011). While many of these studies have involved quasi-experimental matched non-randomised designs – which have a tendency to inflate effect sizes (Wilson & Lipsey, 2007) – there have been two major cluster randomised trials conducted in the US. The first of these involved 41 schools and found positive but more modest effect sizes (0.2-0.3) (Borman et al., 2007) than the pooled average effect of +0.5 reported by Slavin et al. (2011). Similarly, the second more recent trial (Quint et al., 2015) involved 37 schools in total and reported even more modest effects, between 0.03 and 0.15 on the subscales of the Woodcock Johnson reading skills measure.

Since its introduction to the UK in 1997, a number of smaller scale studies have been conducted in the UK, which have shown positive effects in the English context (Slavin et al., 2005; Harris et al., 2001) although another evaluation found mixed effects (Tymms & Merrell, 2001). The most recent largescale quasi-experimental evaluation in the UK involved 40 schools and also found mixed effects, reporting a positive effect of the programme (ES=0.20-0.25) on some school-level literacy outcomes (word level and decoding skills) and no effect on reading comprehension or fluency (Tracey et al., 2014). While there is a substantial body of existing evidence of the effectiveness of SfA in both the US and the UK, it has only been evaluated through two large-scale RCTs in the US and has not yet been robustly evaluated in the UK. SfA is identified as a 'promising' programme on the Blueprints Program website (**http://www.blueprintsprograms.com**) and was also identified as one of the 25 best Early Intervention programmes in the UK in the Allen report (Allen, 2011). There is now sufficient evidence of promise that there is a clear need for a high quality rigorous independent randomised controlled trial to strengthen the evidence behind the programme and its implementation in the UK context.

Evaluation objectives

This is a report on an effectiveness trial of the Success for All (SfA) programme. The evaluation took place between September 2013 and June 2016 and its purpose was to determine the impact of SfA on the literacy outcomes of Reception class pupils and explore issues related to implementation and delivery. The evaluation consisted of two components:

1. An impact evaluation, which aimed to:

    - Assess the impact of the intervention on the literacy outcomes of Reception class pupils at mid-point through the intervention, i.e. at the end of Reception year after 12 months of receiving the intervention;
    - Assess the impact of the intervention on the literacy outcomes of Reception class pupils at the *end* of the intervention, i.e. at the end of Year 1 after 24 months of receiving the intervention;
    - Determine the impact of the intervention for different groups of students, specifically, whether the impact of SfA varied according to: socioeconomic/FSM status and baseline attainment.

2. A process evaluation, which aimed to explore whether SfA was delivered with fidelity, and to identify the successes and challenges associated with implementation.

Ethical review

Ethical approval was granted by the Department of Education Research Ethics Committee, University of York in April 2013, and by the Research Ethics Committee, School of Education, Queen's University Belfast in February 2014, prior to any data collection being undertaken. Informed consent was obtained at the school level from the headteacher, and by individual teachers for interviews and classroom observations. Written information and opt-out consent forms were sent home to parents of eligible pupils (see Appendix C).

Project team

The independent evaluation team was led by Dr Sarah Miller and Dr Andy Biggart (CESI, Queen's University Belfast) and included Dr Seaneen Sloan, and Dr Liam O'Hare. The evaluation team was responsible for the design, randomisation, analysis, and reporting of the results of the impact evaluation in addition to the collection and analysis of the process evaluation data.

The project delivery team was led by Professor Bette Chambers and Dr Louise Tracey from the Institute for Effective Education (IEE) at the University of York and included Louise Elliott and Kate Thorley. The project delivery team worked alongside the Success for All Foundation (Joanne Maxwell) to recruit schools and undertake data collection for the trial at both the immediate and 12-month post-testing points.

Trial registration

The trial was registered with the ISRCTN registry (ref: ISRCTN97011396).

Methods

Trial design

The evaluation was designed as an effectiveness trial to test whether SfA can work at scale. There were two trial arms: control and intervention. As the programme is designed to be delivered across a whole school, randomisation was at the school level, and the programme was delivered with Reception class pupils (across the whole year group) in schools that were randomly allocated to the intervention group. Reception class pupils in the control group were taught the curriculum using standard practice and control schools were paid £2000 per year for their participation. This design was chosen to provide a robust evaluation of the SfA programme with a low risk of contamination between trial arms.

It was originally intended that 50 schools would be recruited to the trial (see 'Sample size' section for more details); however, some difficulties were experienced with school recruitment and attrition in the early stages of the trial, mainly due to the level of commitment required from schools. An initial batch of 39 schools was randomised in July 2013 (with pre-testing to take place at the beginning of the 2013/14 school year). Following some early drop-out of intervention schools (n=7) and the initial difficulties in recruiting the number of target schools, a second cohort of schools (n=14) was recruited the following year. This second cohort of schools was randomised in May 2014 and pre-tested at the beginning of the 2014/15 school year. All schools in both cohorts were followed up for a two-year period on an intent-to-treat basis. Some of the withdrawn schools continued to implement the programme but this varied between schools.

Participant selection

Schools throughout England (but primarily in the North and the Midlands) were recruited by the project delivery team with the support of the Success for All Foundation. The recruitment strategy was multi-stranded and included writing to potentially eligible schools, approaching local authorities and academy trusts, and advertising on the SfA website. Due to the substantial commitment required from schools to participate in the programme, some schools expressed reluctance to take part.

Schools were targeted on the basis of a higher than average proportion of pupils receiving free school meals and the following eligibility criteria were applied:

- Willing to be randomly assigned to condition at the school level
- Willing to engage with the intervention and implement it with Reception classes over a two-year period (until the end of Year 1)
- Willing to allow access to classes for administration of the pre- and post-tests.

All children in Reception class of participating schools were eligible to take part. School-level consent from the headteacher was sought prior to randomisation. Opt-out parental consent for participating in the trial was also sought through schools (see Appendix C).

Outcome measures

Primary outcome: The primary outcome was literacy ability, measured by the Woodcock Reading Mastery Test III (WRMT III) (Woodcock, 2011). Outcomes were collected at mid-point (end of Reception class), using the Letter Identification, Word Identification and Word Attack subscales of the WRMT III. Post-testing was conducted at the end of Year 1, using the Word Identification, Word Attack and Passage Comprehension subscales. For analysis, the relevant subscales were combined to create an overall score representing literacy ability.

Secondary outcome: Phonics Check (standardised national school literacy assessment, administered with all pupils in Year 1 in June 2015 for cohort 1 and June 2016 for cohort 2) was accessed through the National Pupil Database and used as an additional outcome.

Pre-test: The British Picture Vocabulary Scale (BPVS) (Dunn et al., 1997) scale was administered at the beginning of the school year (September/October 2013 for cohort 1 and September/October 2014 for cohort 2) and acted as a pre-test measure of prior pupil attainment in analysis. Although a measure of receptive vocabulary rather than literacy, the BPVS was chosen as a suitable pre-test measure because it has been shown to correlate with later literacy achievement and its use had the potential to facilitate comparisons with other similar trials which used the BPVS as a baseline measure of attainment (e.g. Tracey et al., 2014).

Both the pre-test (BPVS) and post-test (WRMT III) administrations were conducted by trained fieldworkers who were recruited and trained by IEE. Fieldworkers carried out the assessments in school on a one to one basis with each participating child and the tests were scored by the IEE project team. It is possible that because the data collectors were part of the IEE team and thus not entirely independent, this might pose a threat to the validity and independence of the data. However, this threat is diminished by ensuring that the fieldworkers were blind to allocation and following test administration, papers were sent to CESI for data entry and analysis. Table 1 summarises the tests used and the timing of the administration of each.

Table 1: Summary of outcome measures and timing of administration

Date	Timing of outcome data collection		
	Start of Reception Pre-test	End of Reception Mid-point	End of Year 1 Post-test
Pre-test			
BPVS	✓		
Primary outcome			
WRMT III subscale			
Letter Identification		✓	
Word Identification		✓	✓
Word Attack		✓	✓
Passage Comprehension			✓
Secondary outcome			
Phonics Check			✓

It was originally intended that data on attention deficit related behaviours would be collected as a secondary outcome. However, as described in the protocol, this outcome was subsequently dropped and data was not collected because it was determined to be primarily of academic interest and thus not funded by EEF.

Sample size

A sample size calculation was conducted using Optimal Design software (http://hlmsoft.net.od/). It indicated that 50 schools (approximately 25 pupils per school) would be required for the trial to be sufficiently powered to detect an effect of 0.22 of a standard deviation. The following parameters were used in the calculation:

- Significance level (α) = 0.05
- Power (P) = 80%
- Site size (n) = 25
- Intra-Cluster Correlation (ρ) = 0.10
- Proportion of variation at level 2 (R2l2) = 0.60

Randomisation

The unit of randomisation was the school and allocation was conducted independently by the evaluation team once schools had been recruited by Success for All (UK) and the IEE. The randomisation process was undertaken by first ranking schools in terms of the proportion of pupils above Level 4 in English and Maths at Key Stage 2 (averaged over three years where possible) and then randomly assigning each school a random number between zero and one (using decimal fractions to 15 decimal places). Schools were then allocated by the evaluation team (QUB) in pairs with the higher random number of the pair being allocated to SfA and the other to control.

Analysis

Analysis was conducted on an intention-to-treat basis on a combined dataset of the two cohorts. The initial characteristics of the intervention and control groups were compared at baseline in relation to their core characteristics: gender, FSM eligibility, mean scores on the pre-test, and school attainment.

The main effects of the intervention were estimated using multilevel modelling to take account of the clustered nature of the data (where pupil is level 1 and school is level 2). Firstly, a simple analysis was conducted: the relevant outcome measure at post-test forming the dependent variable and the independent variables including a dummy variable representing whether the pupil was a member of the intervention or control group (coded ‘1’ and ‘0’ respectively) and pupils’ baseline scores on the pre-test (BPVS). Three level models were also run to take account of the paired random allocation; however there was no evidence of clustering at level three (the pair) and the models converged with varying degrees of success (and not for the primary outcome). When models did converge there was very little change in the parameter estimates and for these reasons, two level models were conducted and are reported below.

It was intended that a series of pupil-level and school-level characteristics would be added as covariates to control for any baseline differences in the variables; however, since there were no baseline differences between groups (see ‘Pupil characteristics’ section below) the only other covariate that was included in these additional models was school-level attainment at KS2. This covariate was included because it was used in the allocation process and there were baseline differences between the groups on this variable.

The main focus for the analysis was the estimated coefficient associated with the dummy variable that represents the difference in mean scores on the respective outcome variables between the intervention and control groups, once baseline scores and other covariates are controlled for. These coefficients were then used to estimate the effect size of the programme in relation to the respective outcome variables as the standardised mean differences between the two groups at post-test (Hedges' g).

To estimate the effect of the intervention for children eligible for FSM an interaction term between group allocation and FSM eligibility was created and included as a covariate in the analysis. In addition, the main analysis was repeated on a subsample of the students that were identified as eligible for FSM. An additional subgroup analysis examined the differential response to the intervention according to different abilities at baseline (BPVS). Further exploratory analysis examined variation in outcomes according to fidelity of implementation.

Since the proportion of missing data was greater than 5% – the cut-off specified in the protocol – multiple imputation was used and is reported as a sensitivity analysis. This is described in more detail in the results section.

Implementation and process evaluation

The process evaluation involved lesson observations, interviews with staff, and an online teacher survey. All intervention schools were invited to take part in the process evaluation. Classroom observations were undertaken in those intervention schools that agreed to take part (n=11).

Interviews (ranging in length from 20 to 50 minutes) were conducted by the evaluation team with 12 teachers from ten intervention schools. The implementation survey was completed by 20 teaching staff in ten schools. In addition, implementation fidelity was rated by SfA in 15 of the 20 intervention schools still delivering the programme. Each school's implementation was classified into one of three stages: mechanical, routine, or refined (where mechanical is the lowest rating and refined is the highest rating and indicative of high quality implementation fidelity).

Costs

Cost data was collected directly from SFA (UK). Data was collected on the cost of training and ongoing support as well as programme materials, which were estimated based on costs to a school with one-form entry. Cost per pupil was calculated by dividing the cost per school by an average school size of 275 pupils (based on DfE school census statistics for England 2015–2016).

Timeline

Table 2: Timeline

Date	Activity
Evaluation planning	
March – August 2013	Evaluation design (CESI/IEE)
March – June 2013 (cohort 1) **March – June 2014 (cohort 2)**	Recruitment of schools (IEE)
April 2013 (IEE) **February 2014 (CESI)**	Ethical approval (IEE & CESI)

May/June 2013 (cohort 1) **May/June 2014 (cohort 2)**	Randomisation of schools (CESI)
Impact evaluation	
September/October 2013 (cohort 1) **September/October 2014 (cohort 2)**	Administration of pre-test (IEE)
September 13 – June 15 (cohort 1) **September 14 – June 16 (cohort 2)**	Intervention delivery: two academic years for each cohort (Success for All Foundation)
May/June 2014 **May/June 2015** **May/June 2016**	Cohort 1 mid-point test (Administration by IEE) Cohort 1 post-test and cohort 2 mid-point test Cohort 2 post-test
January/August 2016	Collection of KS results (IEE)
November/December 2013 **July/August 2014** **January/February 2015** **September/October 2015** **May/June 2016**	Data entry (CESI)
January 2017	Final analysis and report (CESI)
Process evaluation	
November/December 2013	Process design (CESI)
March/April 2014 (cohort 1) **March/April 2016 (cohort 2)**	Survey administration (CESI)
May/June 2015 (cohort 1) **March/April 2016 (cohort 2)**	Telephone follow-up (CESI)
November 2016	Survey analysis (CESI)
January - April 2014	In-depth interviews and observation (CESI)

Impact evaluation

Participants

Figure 1: Participant flow diagram

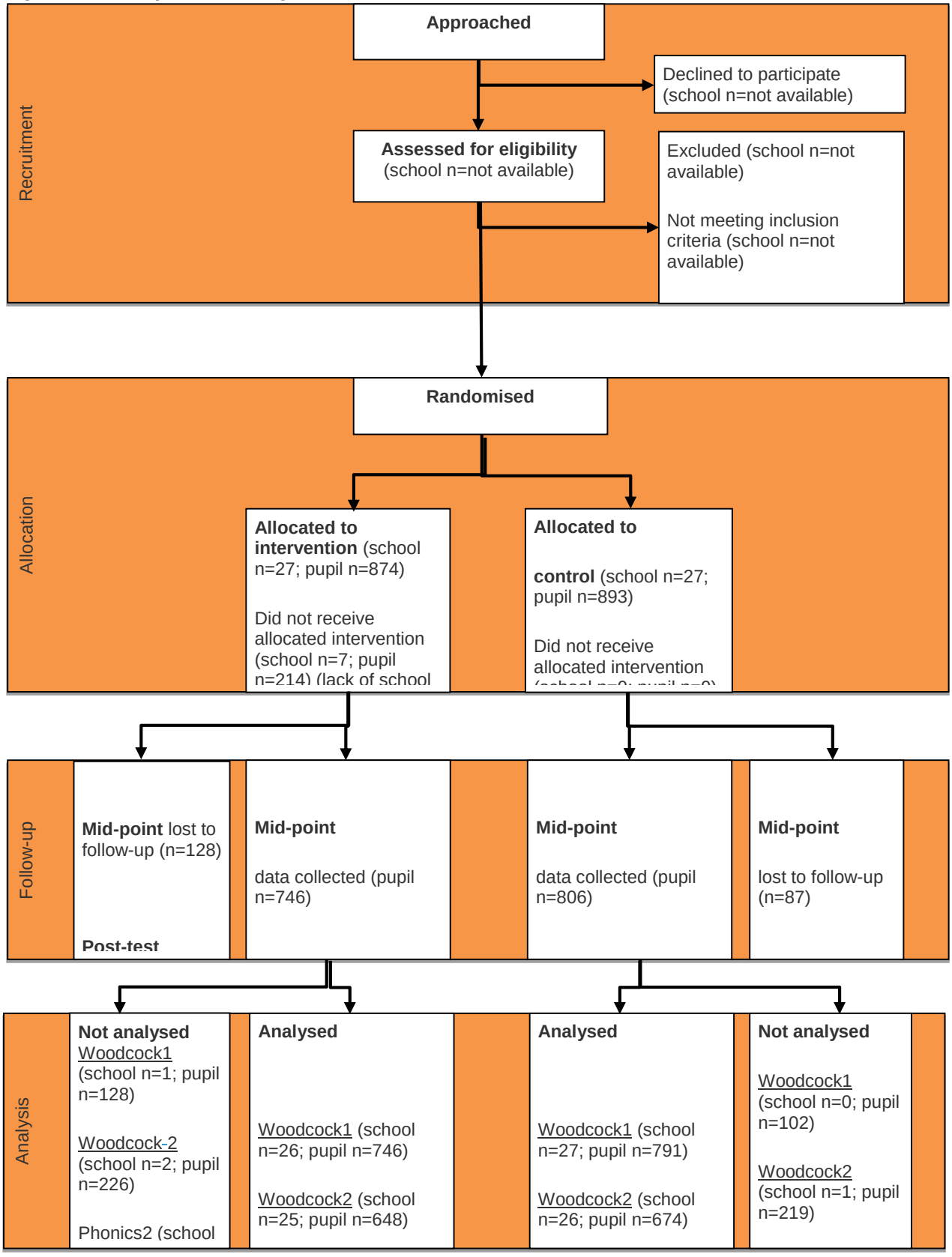

Table 3: Minimum detectable effect size at different stages

Stage	N [schools] (n=intervention; n=control)	Correlation between pre-test (+other covariates) & primary outcome at mid point	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)
Protocol	50 (25:25)	0.60	0.10	Pair matching	80%	0.05	0.22
Randomisation	54 (27:27)	0.20	0.14	Pair matching	80%	0.05	0.31
Analysis	52 (26:26)	0.20	0.13	Pair matching	80%	0.05	0.31

The power of the trial to detect an effect of 0.2 was compromised due to the lower than expected relationship between the primary outcome and the covariates. On reflection, a measure of early literacy might have been an appropriate additional pre-test variable to include and may well have resulted in a stronger relationship (than the BPVS) with the Woodcock Reading Mastery Test, which was used at the mid-point and at post-test. This lack of relationship between the pre- and post-test measures means that the explanatory power of the covariates has been diminished.

Overall, 54 schools took part in the trial and 27 were allocated to each of the intervention and control groups. Post randomisation, seven schools withdrew from taking part in the SfA programme and six remained in the trial on an intention to treat basis. All seven withdrawn schools were in the intervention group and compared to the trial schools, those that withdrew tended to be Academies, have smaller pupil enrolments and fewer children eligible for FSM or with EAL status.

The trial schools were broadly representative of the general population of primary schools in terms of school type and Ofsted ratings. Trial schools tended to be smaller than the national average and – consistent with the requirements for an EEF trial – contained a higher proportion of children eligible for FSM (22% compared to 15%). Trial schools also had slightly lower school-level attainment (72% compared to 80%) and a smaller proportion of children with EAL status (15% vs 20%) and SEN (1.3% vs 2.6%).

Table 4: Characteristics of participating and withdrawn schools compared to the national average

School-level characteristics	National average	Trial schools (n=47)	Withdrawn schools (n=7)
	n (%)	**n (%)**	**n (%)**
School type			
Academy (Converter and sponsor-led)	3008 (18%)	8 (17%)	3 (43%)
Local Authority Maintained Schools	13770 (82%)	39 (83%)	4 (57%)
Community school	*7792 (46%)*	*19 (40%)*	*1 (14%)*
Foundation school	*693 (4%)*	*6 (13%)*	*1 (14%)*
Voluntary aided school	*3133 (19%)*	*9 (19%)*	*0*
Voluntary controlled	*2152 (13%)*	*5 (11%)*	*2 (29%)*
Ofsted rating			
Outstanding	3014 (19%)	7 (15%)	0
Good	11528 (71%)	33 (70%)	6 (86%)
Requires improvement	1435 (9%)	3 (6%)	1 (14%)
Inadequate	153 (1%)	1 (2%)	0
No Ofsted assessment	-	3 (6%)	0
	Mean	**Mean**	**Mean**
School size Total number of pupils	260	238	202
Free School Meal eligibility Proportion of eligible pupils in primary schools	15%	22%	12%
School attainment KS2 English and Maths (both)	80%	72%	71%
English as an Additional Language Percentage of pupils with EAL	20%	15%	12%
Special Educational Needs Percentage of pupils with statement of SEN or EHC plan	2.6%	1.3%	1.5%

Pupil characteristics

In total 1767 pupils with a mean age of 4 years and 6.5 months took part in the trial. Of these pupils, 22% were 'ever' eligible for FSM and 49% were male. Twenty-seven schools were allocated to each of the intervention and control groups (n=54 schools in total). Within this, n=874 pupils were in the intervention group and n=893 pupils were in the control group.

The intervention and control groups were broadly comparable across all the school and pupil level baseline characteristics that were measured (see Table 5). The control schools had a slightly higher

level of attainment (mean=72.8, SD=10.8) compared to the intervention schools (mean 71.5, SD=12.2), which equates to an effect size of 0.11 (Hedges' g) and is statistically significant (p=0.02).

With respect to the pupil-level variables, the control group had a slightly higher proportion of pupils ever eligible for FSM (22%) compared to the intervention group (21%) and a higher proportion of males (49%) compared to the intervention group (48%). In both cases, this equates to an effect size of 0.04. In addition, the control group scored slightly lower on the pre-test measure of vocabulary (mean=59.8, SD=20.6) compared to the intervention group (mean=61.7, SD=18.3), representing an effect size of 0.1 (Hedges' g). Multilevel models were used to compare the groups across all the pupil-level baseline characteristics and no significant differences were found between groups. This suggests that the allocation process worked to create two equivalent groups at baseline.

Table 5: Baseline comparison

Variable	Intervention group		Control group	
School-level (continuous)	**n (missing)**	**Mean (SD)**	**n (missing)**	**Mean (SD)**
School attainment (KS2 results)	27 (0)	71.5 (12.2)	27(0)	72.8 (10.8)
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Ever eligible for FSM	186/874 (114)	21%	186/893 (135)	22%
Males	417/874 (8)	48%	434/893 (25)	49%
Pupil-level (continuous)	**n (missing)**	**Mean (SD)**	**n (missing)**	**Mean (SD)**
Pre-test score (BPVS raw score)	873 (1)	61.7 (18.3)	874 (19)	59.8 (20.6)
Age (months)	869 (5)	54.5 (3.6)	852 (41)	54.5 (3.9)

Outcomes and analysis

Main analysis

The main analysis of the primary and secondary outcomes included only two independent variables: the pre-test measure of vocabulary and group allocation. There was no evidence to suggest that the intervention group had better literacy skills compared to the control group on the Woodcock at the mid-point (Hedges' g=0.04, p=0.44). At post-test (end of Year 1) there was a small positive effect of the programme on the intervention group on both the Woodcock (Hedges' g=0.07, p=0.14) and the phonics check (Hedges' g=0.06, p=0.20); however neither effect was statistically significant.

Table 6: Main analysis using only pre-test vocabulary (BPVS) as a covariate

	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]	p-value
Primary outcome							
Woodcock total score at mid point (end of Reception class)	746 (128)	57.7 [55.7, 59.6]	791 (102)	54.8 (53.0, 56.7)	1537 (746, 791)	0.04 [-0.06, 0.14]	0.42
Woodcock total score at post-test (end of Yr1)	648 (226)	82.3 [79.7, 84.9]	674 (219)	77.6 (75.0, 80.1)	1322 (648, 674)	0.07 [-0.03, 0.18]	0.14
Secondary outcome							
Phonics mark at post-test (end of Yr1)	753 (121)	34.7 [34.1, 35.2]	738 (155)	33.4 (32.8, 34.0)	1491 (753, 738)	0.06 [-0.04, 0.16]	0.21

For information, the raw post-test means of the Woodcock subscales are reported in Table 7. However, to avoid multiple testing and because the global score reported in Table 6 showed no evidence of improvement above that of the control group, the group differences in these subscale means were not tested.

Table 7: Raw post-test mean scores for the subscales of the Woodcock Reading Mastery Test

Woodcock subscale	Raw means			
	Intervention group		Control group	
	n	Mean (SD)	n	Mean (SD)
End of Reception class				
Letter identification	754	31.5 (7.3)	807	30.8 (7.7)
Word identification	754	16.7 (14.2)	807	15.6 (14.0)
Word attack	747	9.4 (8.4)	806	9.1 (8.2)
End of Year 1				
Word identification	650	42.6 (17.1)	691	40.4 (17.1)
Word attack	650	22.6 (9.3)	690	21.8 (9.3)
Passage comprehension	649	17.0 (9.1)	688	15.6 (9.0)

As per the analysis protocol described previously, school attainment was included as an additional covariate in the above model. This was fully observed for all cases and made little or no difference to the magnitude or significance of the estimated effect when included in the model (see Table 8).

Table 8: Main analysis using pre-test vocabulary (BPVS) and school attainment as covariates

	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]	p-value
Primary outcome							
Woodcock total score at mid-point (end of Reception class)	746 (128)	57.7 [55.7, 59.6]	791 (102)	54.8 (53.0, 56.7)	1537 (746, 791)	0.04 [-0.06, 0.14]	0.39
Woodcock total score at post-test (end of Yr1)	648 (226)	82.3 [79.7, 84.9]	674 (219)	77.6 (75.0, 80.1)	1322 (648, 674)	0.09 [-0.02, 0.20]	0.14
Secondary outcome							
Phonics mark at post-test (end of Yr1)	753 (121)	34.7 [34.1, 35.2]	738 (155)	33.4 (32.8, 34.0)	1491 (753, 738)	0.06 [-0.04, 0.16]	0.19

Subgroup analysis

Two pre-specified subgroup analyses were conducted to determine whether the intervention worked better or differently for different subgroups of children.

Firstly, an interaction term between group allocation and (ever) FSM eligibility was created and included in the main model. There was no evidence of a significant interaction in the model that explored effects on literacy (Woodcock) at the end of Reception class (p=0.21) or at the end of Year 1 (p=0.90). There was however some evidence of a possible interaction effect on phonics at the end of Year 1 (p=0.08). For this reason, the main analysis reported above – which included only the pre-test score and group allocation as independent variables – was repeated with only those pupils identified as ever having been eligible for free school meals. This resulted in an overall subsample of n=386 pupils (although within this there was some missing data) and the results of this analysis are reported in Table 9. It appears that there is some evidence that the SfA programme benefits children eligible for FSM, particularly with respect to mid-test literacy ability as measured by Woodcock at the end of Reception class (Hedges' g=0.22, p=0.03) and the post-test phonics check at the end of Year 1 (Hedges' g=0.16, p=0.06). While FSM pupils in the intervention group also scored better than their peers in the control group on the Woodcock literacy test administered at post-test (the end of Year 1), this result was not statistically significant.

Table 9: Main analysis using only pre-test vocabulary (BPVS) for pupils eligible for FSM

	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]	p-value
Primary outcome							
Woodcock total score at mid-point (end of Reception class)	164 (22)	51.71 [47.51, 55.92]	174 (26)	44.90 [41.13, 48.66]	338 (164, 174)	0.22 [0.01, 0.44]	0.03
Woodcock total score at post-test (end of Yr1)	153 (33)	69.97 [64.27, 75.68]	168 (32)	64.46 [59.06, 69.87]	321 (153, 168)	0.12 [-0.10, 0.34]	0.23
Secondary outcome							
Phonics mark at post-test (end of Yr1)	185 (1)	32.43 [31.08, 33.77]	191 (9)	30.04 [28.40, 31.68]	376 (185, 191)	0.16 [-0.04, 0.36]	0.06

To test whether the intervention worked better for pupils who had low pre-test scores, the age-standardised BPVS scores were categorised into two groups: scores less than 85 (i.e. less than one standard deviation below the mean n=251) and scores 85 and above (n=1468). An interaction term was created between group allocation and this dichotomised version of pre-test attainment and included in the main model alongside group allocation and pre-test scores. The resulting models did not provide any evidence of an interaction in relation to any of the three outcomes, suggesting that the SfA programme does not result in different outcomes for children with different levels of ability at baseline. However, given that there was not a strong relationship between the baseline and post-test measures it may well be the case that children with lower levels of literacy or phonological skills (rather than lower levels of vocabulary) might have benefitted more from the programme.

Table 10: Sub group analysis using only pre-test vocabulary (BPVS) for pupils with low baseline attainment

		Raw means					
		Intervention group		Control group		Effect size	
Outcome	Baseline attainment	n (missing)	Mean [95% CI]	n (missing)	Mean [95% CI]	Hedges' g [95% CI]	p-value of interaction term
Primary outcome							
Woodcock total score at mid-point (end of Reception class)	Not low baseline attainment	655 (102)	60.11 [58.06, 62.17]	654 (57)	58.35 [56.33, 60.37]	0.02 [-0.26, 0.29]	0.55
	Low baseline attainment	90 (22)	39.51 [34.49, 44.53]	121 (18)	36.20 [32.18, 40.22]	0.05 [-0.06, 0.16]	
Woodcock total score at post-test (end of Yr1)	Not low baseline attainment	574 (183)	86.10 [83.45, 88.71]	567 (114)	81.35 [78.66, 84.04]	-0.06 [-0.36, 0.23]	0.16
	Low baseline attainment	74 (38)	52.70 [44.83, 60.58]	102 (37)	56.05 [49.33, 62.76]	0.12 [-0.001, 0.23]	
Secondary outcome							
Phonics mark at post-test (end of Yr1)	Not low baseline attainment	666 (91)	35.50 [35.03, 35.97]	630 (81)	34.44 [33.88, 34.99]	0.06 [-0.22, 0.35]	0.84
	Low baseline attainment	87 (25)	28.34 [25.72, 30.97]	102 (37)	27.11 [24.55, 29.66]	0.10 -0.01, 0.21][	

On treatment analysis

Given that seven intervention schools withdrew at the start of the trial, an analysis was conducted to explore the consequence of running the main models again, but this time excluding the data from the withdrawn schools. As can be seen from Table 11, when discounting the data from the withdrawn schools there is a small positive effect in favour of the intervention group at mid-point (Hedges' g=0.06, p=0.19) and for the phonics check at post-test (Hedges' g=0.07, p=0.22) but these findings are not statistically significant. There is however some evidence of a small statistically significant positive effect in favour of the intervention group on the Woodcock literacy test at post-test (Hedges' g=0.10, p=0.05).

Fidelity was rated for 15 of the intervention schools whereby implementation was classified as mechanical, routine, or refined, or at mid-points between these levels (see Appendix D for further details of how implementation fidelity was rated). For the purpose of analysis, schools that were rated as refined or refined/routine were coded as 1 (n=7) and schools rated as either routine or mechanical were coded as 0 (n=8). This dummy variable for fidelity was then included in an analysis of the intervention

data only, to explore whether fidelity was associated with a better outcome on any of the three literacy measures. There was no evidence that fidelity affected Woodcock scores at the end of Reception class (Hedges' g=.11, 95% CI [-.08, .30]); however, there was some evidence that schools rated as 'refined' resulted in higher scores on both the Woodcock (Hedges' g=.20, 95% CI [-.01. .40]) and particularly the phonics check at post-test (Hedges' g=.41, 95% CI [.21, .61]), i.e. the end of Year 1. This analysis should be treated with caution due to the smaller samples involved and potential bias given that fidelity (as used in this analysis) was rated by SfA and not independently.

Table 11: Main analysis excluding withdrawn schools and using only pre-test vocabulary (BPVS) as a covariate

	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]	p-value
Primary outcome							
Woodcock total score at mid-point (end of Reception class)	598 (276)	58.19 [56.00, 60.38]	791 (102)	54.83 [52.96, 56.70]	1389 (598, 791)	0.06 [-0.05, 0.17]	0.19
Woodcock total score at post-test (end of Yr1)	547 (327)	82.65 [79.78, 85.53]	647 (246)	77.55 [74.98, 80.13]	1221 (547, 647)	0.10 [-0.01, 0.22]	0.05
Secondary outcome							
Phonics mark at post-test (end of Yr1)	591 (283)	34.55 [33.93, 35.17]	738 (155)	33.41 [32.79, 34.03]	1329 (591, 738)	0.07 [-0.04. 0.17]	0.22

Sensitivity analysis: multiple imputation

The proportion of missing data at both post-test and the 12-month follow-up for all outcomes was greater than 5% (see Table 12), indicating that a re-analysis of the data using multiple imputation would be appropriate. There was a higher proportion of missing data (15%) in the intervention group compared to the control group (10%) for the immediate post-test outcome and this difference was statistically significant (Chi2 (1) =9.5, p=0.002). For both of the 12-month post-test outcomes the proportion of missing data was not significantly different between the intervention and control groups.

Missing-ness at post-test was also statistically significantly associated with poorer scores on the pre-test variable across all post-test outcomes, including: immediate post-test literacy ability (p=0.003), 12-month post-test literacy ability (p=0.002), and 12-month post-test phonics check (p<0.001).

Table 12: Proportion of missing data for each outcome

Outcome	Intervention missing n (%)	Control missing n (%)	Total missing n (%)
Pre-test receptive vocabulary (BPVS)	1 (0.1%)	19 (2%)	20 (1%)
Mid-point (end of Reception class) literacy ability (Woodcock)	127 (15%)	87 (10%)	214 (12%)
Post-test (end of Yr1) literacy ability (Woodcock)	225 (26%)	205 (23%)	430 (24%)
Post-test (end of Yr1) phonics check (SAT)	120 (14%)	141 (16%)	261 (15%)

The imputation model imputed the data separately for the intervention and control groups and included all relevant variables and auxiliary variables involved in the analysis and sampling design. Data was presumed to be missing at random (MAR) rather than MCAR (missing completely at random) or MNAR (missing not at random). This assumption renders the missing mechanism ignorable, simplifying the imputation step while ensuring correct inference. The imputation was performed using chained equations which fills in missing values in multiple variables iteratively by using a sequence of univariate imputation methods with fully conditional specification of prediction equations. This method accommodates arbitrary missing-value patterns. Twenty imputations were conducted in order to lessen the simulation (Monte Carlo) error.

Table 13 compares the main complete case analysis (reported previously in Table 6) with the results of the same analysis, but this time using instead the imputed datasets. It can be seen that using multiple imputation elicits broadly the same magnitude of effects as a complete case analysis.

Table 13: Main analysis using both complete case analysis and multiple imputation

	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]	p-value
Primary outcome							
Woodcock total score at mid-point (end of Reception class)							
Complete case analysis	746 (128)	57.7 [55.7, 59.6]	791 (102)	54.8 (53.0, 56.7)	1537 (745, 791)	0.04 [-0.06, 0.14]	0.42
Multiple imputation	874 (0)	53.63 [51.79, 55.47]	893 (0)	51.71 [49.83, 53.58]	1767 (874, 893)	0.04 [-0.05, 0.13]	0.33
Woodcock total score at post-test (end of Yr1)							
Complete case analysis	648 (226)	82.3 [79.7, 84.9]	674 (219)	77.6 (75.0, 80.1)	1322 (648, 674)	0.07 [-0.03, 0.18]	0.14
Multiple imputation	874 (0)	76.68 [74.38, 78.98]	893 (0)	70.06 [67.78, 72.35]	1767 (874, 893)	0.08 [-0.02, 0.17]	0.07
Secondary outcomes							
Phonics mark at post-test (end of Yr1)							
Complete case analysis	753 (121)	34.7 [34.1, 35.2]	738 (155)	33.4 (32.8, 34.0)	1491 (753, 738)	0.06 [-0.03, 0.18]	0.21
Multiple imputation	874 (0)	33.64 [33.11, 34.17]	893 (0)	31.61 [31.00, 32.22]	1767 (874, 893)	0.06 [-0.03, 0.15]	0.17

Cost

SfA costs approximately £46,500 per school to implement with the whole school in the first year. This figure includes the following start-up costs:

Table 14: Breakdown of costs associated with implementing SfA in the first year

Item	Detail	Cost
Training	Delivered by 3 SfA trainers to the whole school over 2 days	£6,800
Support days	8 days at £650 per day, allocated across the school year	£5,200
SfA resources	Teacher and pupil resources for all stages of the programme (including teacher manuals, guidance manuals, guided reading books, flash cards, and online resources). Cost to a school with 1-form entry is £14,000 (cost given here is based on average school size of 1.5 forms)	£21,000
Books	Purchased by schools, either from own supplier or through SfA partner. Cost to a school with 1-form entry is £9,000 (cost given here is based on average school size of 1.5 forms)	£13,500
Total		£46,500

Assuming an average school size of 275 pupils, the average cost of the programme would be approximately £169 per pupil in the first year. This cost is, however, likely to be an underestimate of the actual full economic cost of the programme, which would also need to include estimates of the cost of teacher and teaching assistant time. Additional time also needs to be set aside for the member of teaching staff in each school that acts as the overall SfA co-ordinator. These are important to include as they represent significant 'opportunity costs' in that they are resources that could be used on other programmes.

Beyond the first year, schools would have 3–4 support days (at £650 per day) which would be tailored to the needs of the school, with SfA providing training, coaching, co-teaching, holding staff meetings, or training new staff. These support days are allocated at the school's discretion. Schools may need only to replenish supply of books for younger year groups in subsequent years, at a cost of £15 for 25 books (or £22.50 based on 1.5 form entry). Thus the estimated cost of the programme in subsequent years is £2297.50, comprising:

- Books for Reception class pupils: £22.50
- Support days (an average of 3.5 at £650 per day): £2275

- This means that in subsequent years, the cost of SfA per pupil is £8.35.

Table 15: Cost of SfA per year over three years

Number of years using the programme	Cumulative cost per pupil (£)	Average cost per pupil per year (cumulative cost per pupil/number of years) (£)
1 year	169	169
2 years	177.35	88.68
3 years	185.70	61.90

There was no charge to the schools involved in the current evaluation.

Process evaluation

This section presents the key findings from the process evaluation, bringing together the findings from the various components of the process evaluation, i.e. the lesson observations, the interviews with staff, and the teacher online survey. Of the ten schools that took part in *both* the survey and the interviews, eight had an Ofsted rating of 'good', one was rated 'requires improvement', one was rated as 'inadequate', and the mean proportion of pupils eligible for FSM was 26%.

Implementation

Training and support for delivery

Of those who responded to the survey (n=20), most (90%) felt both suitably qualified and confident to deliver the programme on the basis of the training they received. Teachers were also asked about training and support during the interviews. When asked if the training provided adequate preparation for delivery, one teacher reflected that:

Yes, I think it did. I think while we were actually sitting there actually having the training, it was hard to visualise what it would be like, and it was actually a case of, doing it was the way to really learn it and it was sort of a bit difficult at first – you had your script in one hand and you were trying to follow where you were, but, yeah, it's fine now.

Several teachers commented that it would have been helpful to have visited a school (or specifically, another reception class) that was delivering the programme, to allow them to see it in practice.

> *That was one thing we would have liked to have done, see how it had worked in another school. We had, there were some people came from a junior school near us who had been doing it but of course they're only a junior school so we didn't hear how it worked in Reception. It probably would have given us a better idea, because it took us a while to sort out a timetable to begin with, to fit everything in.*

A large majority (90%) of survey respondents reported that they received sufficient ongoing support; however, this was not always the case during the interviews. One school felt they did not receive enough support:

I just think that we haven't had much at all, really… Well, I certainly have no contact with [SFA] but I know [school SFA coordinator] kind of our leader so she may have done, but not much I don't think

However, this was in contrast to other teachers who felt the ongoing support was very helpful:

[The ongoing support] is really, really helpful, definitely, and the feedback that [SFA consultant] has given us and the targets that we've got to reach, that's been really helpful.

They've come in several times since, and they've observed. [SfA consultant]'s done a whole day's observation here, in all the different classes that are teaching it. She's gone through the whole lot and we've had feedback from that, and just continual email contact about any queries we have as well. So it's quite well supported.

Acceptance of the programme

There was a mixture of excitement and scepticism reported by teachers when the programme was first introduced to the school, as reflected in the following comments from teachers:

> *I was very excited about it, hearing about having the training, the cooperative learning aspect fits in well with the school mission, the way that we want to be. I like the idea of all the resources, so yeah, I was quite excited about it.*
>
> *To begin with when we did do the training I was a little bit sceptical... it's very American and it's hard when you're very British to then adopt very American approaches to doing things. So yeah, there was a little bit of scepticism with it.*

One teacher reflected that while initial impressions of the programme were relatively negative across the school, these had changed after several months of implementation and upon noticing an improvement in pupil writing ability:

> *Initial impression was, "Oh, no," mostly because it was so scripted, and in foundation stage, you haven't just got a script for literacy, you've got a script that you could use from beginning to the end of the day and I think everybody's initial impression was, "Oh, it's all there in the book. Why do you need a teacher to deliver this?" That was the initial feeling, but that has changed dramatically as we went on… when we've seen the results of the children's writing, I think everybody's changed their opinions on it.*

These reflections are consistent with the very positive responses to the online survey (conducted towards the end of the delivery year) whereby 95% of respondents were enthusiastic about their school adopting the programme and 85% considered it necessary for their pupils. All survey respondents indicated that they would be happy for their school to continue with SfA. This notwithstanding, some teachers felt that the pace of SfA was faster than they were used to in Reception class, and initially took some time to get used to.

You've got to get it all in, you've got to have really pacey lessons, but also you've got to give the chance to the children to actually digest what they're supposed to be listening to and taking it on board, and sometimes it just doesn't give them enough time for that.

Pupil acceptance and engagement

Eighty-five per cent of survey respondents felt that the programme motivated and engaged the children; however, a notable minority (25%) did feel that SfA was not suitable for all children. These mixed views were also reflected in the interviews:

> *There was a lot of sitting down we thought, for some children. Some have coped with it okay, but others you can see they're getting a little bit twitchy, you know, and they want to do something so, you sort of have to hone it down a little bit, or split it up a bit more so there's not so much sitting down.*
>
> *They do enjoy the reading and the shared reading and the partner reading, they really enjoy that.*

Several teachers commented that the class enjoyed and quickly became used to the routine of SfA:

> *What they like is the routine of it, they know exactly how their morning is going to be and you notice it if for any reason we have to alter it, or swap something around, they don't like that, they don't like being changed around, so yeah, it's the routine of it is very good.*
>
> *Once or twice you kind of get [groans] but I think on the whole actually they seem to quite like it, I think, which is good.*

On the plus side the themes have proved popular in engaging the boys – themes like "words and roads" have let us use cars and vehicles with them.

Teachers felt that most children in their class enjoyed the fast-paced nature of the programme, but felt that this doesn't suit all children:

Usually every week there's a different activity going on in there and the children find it quite interesting because they get in there, 'Ooh, what's in there today?' Yeah. So they quite like that because it's changing a lot, we change the room around as well sometimes to facilitate things. It doesn't suit everybody. Some of the children are sort of, 'What's happening now?' You know, so you have to give them a bit of extra support.

Other teachers commented that SFA provision at the Reception stage needed to be better differentiated for different ability levels. Some teachers felt that the programme was not as well suited to the lower and higher ability groups, including children with special educational needs:

It's all pitched at the same level anyway that they suggest in SFA so I kind of have to adapt it to suit my children. It's not suited for more able children at all.

I would say some of them are more able children, it is pitched quite low I think so we do have to extend some of the ideas for our more able.

In terms of meeting the needs of some children, children with particular needs, I think that's really quite tricky, especially with the amount of concentration time that's expected. And actually if they've got a particular issue or need, how well is that going to fit in? So I have reservations about that.

No, it doesn't cater for children with special educational needs at all. I've got two very low ability children, but it's more to do with their speech as well... and obviously I'm using all these lovely big words, but they can't actually form them... they're having problems actually speaking and saying the words correctly, so that's not really any benefit to them to be honest and there's nothing in the programme that will help them and benefit them.

> *I think the Write Away where you model the sentence as a group and then you remember the sentence, then you go away and you write it, and then obviously if you're capable you can write more sentences if you want to do, which our top ones are really, really are flying. My middle group are doing really quite well, it's the really poor ones who are struggling because they struggle to retain that sentence, because I said the thing is it's all to do with children with working independently, isn't it, but really a couple of days a week, they need to have an adult supporting them to do their writing otherwise they're not going to get anywhere with it.*

Programme resources

Ninety-five per cent of those who completed the survey found the materials and resources useful and through the interviews it emerged that the books were particularly well regarded:

> *Firstly, it was the books, fantastic books, really, really good books, and that's been probably the biggest resource for me here. We read a book every single day. We might repeat it during the week, but practically, it is a book each day, and that's been the biggest resource. We also have big books as well, that I use for teaching, and the children have their miniature versions of those as well. So really for us here, it's been the books, has been the major resource. Alphabet lines, as you can see hanging across our classroom there. Red words, green words, bits and pieces like the words for the reading as well, when we're reading books.*

Barriers to delivery

Staffing

One of the challenges to implementing the programme was the logistics around staffing the smaller reading groups. As one teacher said:

I think that has been probably something that has been a bit of a problem. Because we're such a small school, we have to really split our groups down to quite small numbers… if a couple of people were absent, then it would change that. Although we'd be able to accommodate it, that would be when you'd run into difficulties, I think. I suppose it depends how small you want to keep your groups though.

Several teachers noted that some of the content and activities were really only viable if there were enough teaching staff available to provide support:

I must admit when I look at it I do think that a lot of it, it sort of says 'with adult support' but in Reception class you're actually quite lucky to have two people, some Reception classes have only got one.

A lot of the resources that they suggest you use in these labs, you need people sat there with and talking to the children, and I haven't got enough adults to sit in each lab and talk to the children about these things.

Another teacher noted that:

I think you'd need some more staff to implement it as well as it should be.

Space

Several teachers commented that their classroom environment at times was not adequate for full delivery of SfA in reception class. Some schools found it problematic to find spaces for the reading groups within reception class, with some groups having to move to other parts of the school:

> *I think one of the things was just the space… you know, for the Reading and Phonics, just trying to have the space and the room to sort of have a little group here, another group there and someone somewhere else. You know, the logistics of… because they do SfA first thing in the morning and you can't get into the library then or whatever you know, so that's been one of the things.*

Similarly, in response to the survey, 40% of respondents did feel that the room in which the programme was being delivered was not adequate for all the activities.

Prescriptive nature

Some teachers noted some initial apprehension about adopting the programme, due to its prescriptive nature, which was different from what some teachers were used to in Reception class.

> *I think if you come in from the outside and you look at it, it must look really quite strange with people with scripts about them, and yet it is scripted and it's a bit American at times, but you can put your own slant onto it, especially I think at this end as well. So I think you have got a bit more freedom than you probably thought, as long as you're doing those essential scripted parts I think there's room for you to put your own slant onto it as well.*

Some teachers felt the highly structured nature may stifle or limit the amount of child-initiated learning, creativity or free play.

> *Because it's so prescribed in that way that you're kind of missing out on those other elements of the arts and craft and not delivering on the carpet with them any kind of expressive arts with them because SFA doesn't allow you to do that… we're missing out on those other elements in the early years curriculum as well though.*

> *It is very different for us, yes, because it tells you exactly what to do, doesn't it, all the time in the Reception so to speak. So it has been quite different. I don't feel there's the same amount of child choice in it because it's quite… children are told exactly what to do like in the art activities and things like that whereas before we would perhaps say to them we're doing the spring; make something that you like to do with the spring.*

While 90% of survey responders considered the programme instructions to be very clear, similar reservations were expressed with regard to the prescriptive nature of the programme, with 80% reporting that they felt the programme to be too prescriptive. Furthermore, 70% thought that the programme sometimes missed important elements that they felt should have been covered. In interviews, teachers commented that they would like to see other aspects incorporated into the programme, such as outdoor learning opportunities, traditional tales, and better use of ICT and multimedia:

> *Well the big thing I think in early years especially is I'd expect more traditional tales element. I think that's quite important for story-telling language with children... and that's just not really covered.*
>
> *I think probably more outdoor learning opportunities I think would be good.*
>
> *I think the ICT aspect of it and the media side of it is quite poor to be honest, because I mean we have children and they are using like iPads and interactive work boards and all those kinds of things and none of that is in that programme.*

One teacher also felt the Write Away activities could be strengthened:

> *I think that's my issue with the Write Away is that it seems to sometimes be writing a sentence just for writing a sentence. And I know it's important to build up those skills but it's the exciting opportunities for writing for purpose that I think I have to make sure that are included.*

Amount of content

While 75% of those who completed the survey considered the frequency of the programme to be sufficient to impact on children's attainment, the majority of teachers – both interviewed and surveyed (85%) – reported struggling to incorporate all of the content into the school day. While the fast-paced nature of the programme was beneficial in keeping children interested, at times teachers found it difficult to keep up with the demands of the programme as the themes change:

I think the themes are quite nice but I think as we're changing themes every two weeks it seems sometimes it's quite hectic… I think when we did the theme, like the community and then one day we were doing the fire fighters, the next day we were doing the police, the next day we were doing crossing the road, the next day we were doing…, I think it was wearing safety helmets and the next day it was… we did five different things in a week and the children did enjoy it but it's a lot.

In some cases, teachers felt the amount of content was disproportionate to the amount of staff available:

I think there is too much to do, I mean we do the 15 minutes maths, we do the sounding out, we do the reading, we do the writing, and we do the write-away and we do elements of the science. But if we were to do every single thing that was in the book every day, it would be a struggle, and we can't do the snack time because we don't have the staff, we just haven't got the available adults. It was a question of the available adults…

Lack of differentiation

There were mixed opinions about whether the programme was equally suited to, or meeting the needs of, all children in the class. Only 20% of those surveyed considered that a different programme would have been more suitable to meet the children's needs and 75% thought the programme was suitable

regardless of children's needs. The teacher interviews however, did highlight some issues in relation to differentiation:

> *We've had to differentiate some of the activities as well, I mean, some of the things were too hard for the less able children to grasp and then I've got some more able children so we've had to move them on a little bit further because it wasn't enough to challenge them.*
>
> *Also there's no differentiation with what they give you, they just find every child works at the same level but that's not how it is in a classroom because I've got children who are still trying to write their name, because they're my SEN, and I've got children who are exceeding the early learning goals already, they're above and beyond where they should be.*
>
> *It is pitched quite low I think so we do have to extend some of the ideas for our more able.*
>
> *For my SEN children it's absolutely fantastic, it really ticks all the boxes for them, it's been brilliant for them. But also for the ones that need to be pushed that little bit further, it's had the option for that as well, because it's just so easy to differentiate all those questions when you're on the carpet, and with the writing activities as well it's quite open, so you can sort of cater for all of them.*

Fidelity

As part of the intervention, Success for All rated the implementation fidelity of 15 of the 20 intervention schools still implementing SfA at both the end of Year 1 and the end of Year 2. Of the five schools that were not rated, one school's rating was missing and the remaining four were not implementing the programme sufficiently to be awarded a rating and were presumed by Success for All to be withdrawn from the programme. This means that of the 27 schools allocated to the intervention group, only 15 implemented the programme; 12 did not.

Fidelity was rated by SfA using a standardised form provided in Appendix D. There are three classifications for implementation fidelity: mechanical (stage 1), routine (stage 2), and refined (stage 3). Schools are expected to move through each stage until they reach refined status which is the highest level of implementation.

- A rating of ***mechanical*** means that a school is in the initial stage of implementation and whether or not teachers in each classroom are using the basic lesson structure and objectives as well as using the available media regularly and effectively. At this stage, pupils should also be familiar with the routines of the programme.
- A rating of ***routine*** indicates that specific processes and objectives within the programme are being met regularly and there is an increased level of pupil engagement.
- A rating of ***refined*** means that more fine-grained, nuanced processes and objectives are being met by teachers and engagement is high on the part of pupils.

Once individual classrooms have been rated according to the relevant processes and objective of the programme, the school is given a summary classification for each component to reflect the stage of implementation:

- Learning – staff are working towards verification of this objective
- Significant use – objective is verified for 40% of teachers
- Mastery – objective is verified for 80% of teachers
- Power schoolwide – objective is verified for 95% of teachers

Of the 15 schools that were still actively using SfA, at the end of the first year, one school was rated as demonstrating 'mechanical' delivery – the poorest rating on the three-stage scale of: mechanical,

routine and refined (the highest rating); one school was rated as 'mechanical/routine'; and the remaining schools were classified as 'routine' delivery.

At the end of Year 2, only one school reached the highest classification of 'refined delivery', with six schools being classified as 'routine/refined', five as 'routine' and three as 'mechanical/routine' delivery. No schools were rated as 'mechanical' by the end of the two-year implementation period. In the qualitative comments that accompanied the ratings, schools that were classified as routine/refined tended to be those that *engaged with the programme* and were described as *fully embracing* SfA. Similarly, schools whose implementation fidelity was rated as mechanical or mechanical/routine tended to be described as *not engaging* with the programme sufficiently. The schools that were rated as refined or refined/routine in their implementation tended to have a slightly better Ofsted rating, fewer pupils eligible for FSM, and higher school-level attainment overall than schools that were rated as routine or mechanical/routine.

Issues around fidelity were also reflected in the comments from teachers about how they felt they had to adapt the programme to best suit their needs:

We do adapt it to suit our own need I would say, we don't follow it exactly to the letter because sometimes we find some of the activities a bit dull or we think we could do something else that the children are more interested in. We don't use a lot of the photocopiable things in the backs of the resources books because again as a school we don't often use that many worksheet-based things for our early years, it tends to be more hands on and investigative stuff.

When we looked at it originally I said we're going to continue with our practice that's been really good so far and everything that we've done and achieved, but we'll adopt some of the SFA ways of doing it and some of the elements of it.

It is very prescriptive, but we don't do it like that. We've taken the bits that we want and the bits that we like and we're just doing them. I'm not completely following it to the script. It doesn't allow you freedom and creativity.

Outcomes

Many teachers (including 90% of survey responders) felt the programme was having a beneficial impact on their pupils in terms of their literacy:

> *I think the children with their reading are certainly getting more confident now with it. You know, the impact of the reading when they're writing, they're obviously thinking about the letter sounds in their writing and stuff like that. So it is having an impact on their writing.*

> *I've noticed with the reading and the writing, our children come in at a very, very low level, but it's been quite a slow process, but as we're sort of getting into Term 5, really seeing a huge improvement in their reading and their writing, and also quite often in themselves that they've got the confidence to have a go at something and to have a go at writing, so that's been good.*

Several teachers felt that there were benefits for the pupils in terms of their speech, their confidence in speaking out in class, and their ability to work cooperatively with each other:

> *Verbally, they'll get it, you know, and they'll always be involved like having the 'talk to your partner', they're more confident and they're happier to share things with you because they know that they're going to get support if they need it. So that's been really good.*

> *I think it's getting the children to speak more clearly and sentences and things.*

> *The random reporter… they like that now, they know if I'm getting it out what's going to happen next and they're quite happy to talk. There's a lot of talking, I mean, the 'let's get together' time we've split into groups now to make it a bit pacier because it was taking a while to get through it all but we've found children are speaking more now because it's a smaller group so that's helped… One little girl that really wasn't speaking is starting to speak now so she's asking me questions.*

> *Immediately I've felt that once we'd introduced the button jars that I've gone for the cooperative learning standards, I was surprised how quickly that the helpfulness really increased. It was lovely celebrating that helpful nature with children and I really feel that straightaway, I mean it was just within a week or two really there was this atmosphere when you've got carpet time together, that they would be helping each other, whispering to each other to help with the answers. Which I thought was lovely, so nobody feels under pressure.*

There were mixed opinions on the impact of the programme on their role as a teacher, with some finding the programme resulted in less lesson planning and preparation, but others reflecting that, at least in the first year of implementation, it had a negative impact:

> *I think there's probably been slightly less preparation with this, just for the sort of literacy section, most of it is all there, so probably perhaps isn't as much of that.*

> *For the first couple of weeks it was almost like a little honeymoon period for it where it was, wow, this is fantastic, and then after that I think it was the enormity of how much to get through bothered me… a little bit stressful thinking about how we were actually going to manage implementing it all. It's taken a lot of my time up thinking about the practicalities of it… So I had to reinvent the wheel a lot and that's taken up enormous amounts of thought power actually which creates the same stresses doesn't it, when as a teacher you've got a lot of other things to be thinking about.... And of course that then impacts on how staff feel about it in terms of that you're having to change routine so much.*

> *It hasn't made it easier at all. The planning side of it, not whatsoever. We can't use the manuals for our planning, as I've said, because it's not differentiated so we've had to do all our own so we've had to change all our plans that we had before.*

Control group activity

As the trial was designed with randomisation at the school level, the risk of contamination between groups was low, and there was no evidence that any schools in the control group were exposed to SfA. Control schools continued with business as usual which consisted of curriculum delivery as normal. Beyond this, what constituted 'business as usual' was not explored.

Formative findings

Essential ingredients of the programme

The essential ingredients of the programme, from the perspective of teachers, were: cooperative learning, the strong focus on literacy, and the resources provided:

The cooperative learning is the main thing I think really, and sort of being part of a team and all that sort of thing… that you know, everybody's a part of that and you've got to play your part, I think.

I would think the main components are literally embedding the kind of reading and writing environment really within early years and that whole kind of literacy rich environment.

For me I think it's that strong phonics base and phonics application from the Rhyme time which there's a lot more to that than I realised at first, and then the Sound Steps.

That would be the quality books that we get to read. They are really, really fantastic books.

Intervention intensity

Most teachers interviewed felt that there was too much content to be delivered over the course of a full day or week, and as a result, adapted the programme to fit within time and staff constraints. If the programme is to fit within the UK context within Reception class, some of the content may need to be removed or made optional.

How the programme could be improved

As several of the teachers interviewed commented on the value of seeing how the programme is delivered in practice, a selection of video recordings demonstrating each element of the programme may be useful.

Teachers generally perceived there to be greater differentiation within the programme, to cater for both higher and lower ability groups of pupils.

The amount of literacy content within the programme was seen as a strength, although some teachers felt that this may be at the expense of other aspects of the curriculum including outdoor learning, physical education, and religious education. Incorporating ICT and multimedia to make use of technology in the classroom was another recommendation for improvement.

Targeting the right children

As described above and in earlier sections, some teachers felt the programme as a whole was less well suited to the highest and lowest ability groups in the class, and that more work could be done to ensure that the highest ability groups are challenged and at the same time, the lower ability groups are supported.

Conclusion

Key conclusions
1. Children who took part in Success for All (SfA) made 1 additional month's progress, on average, after two years compared to children in other schools. The 3 padlock security rating means that we are moderately confident that this difference was due to SfA.
2. Children eligible for free school meals (FSM) made 2 additional months' progress after two years, compared to FSM children in control schools. The smaller number of FSM pupils in the trial limits the security of this result, though combined with other findings in the report it provides some evidence that SfA does improve literacy ability for children eligible for free school meals.
3. Of the 27 schools receiving SfA, 7 dropped out of the programme completely, and 5 did not implement it to the minimum expected level. There is some indication that lack of engagement with the programme resulted in poorer implementation which may have affected the programme's impact.
4. Some schools found SfA prescriptive and adapted the content and delivery of the programme. However, there is some evidence that schools delivering SfA as prescribed by the developers saw larger gains in children's literacy after two years than those that completed fewer aspects of the programme.
5. Overall, schools that successfully delivered SfA were enthusiastic and valued the programme in their school.

Interpretation

The current trial hypothesised that the Success for All programme would improve the literacy outcomes of Reception class pupils after one year of delivering SfA (mid-point through the trial, at the end of Reception class) and these effects would be sustained and evident after two years of delivering the programme (post-test, at the end of Year 1). It was also hypothesised that pupils who were from more deprived backgrounds or who had lower baseline attainment would benefit more from the programme than other children.

The main analysis (reported in Table 6) found no evidence to suggest that the intervention group had better literacy skills compared to the control group at the end of Reception class (Hedges' g=0.04, p=0.42). At post-test (end of Year 1) there was some evidence of a small positive effect of the programme on the intervention group on both the Woodcock (Hedges' g=0.07, p=0.14) and the phonics check (Hedges' g=0.06, p=0.21), although neither of these results is statistically significant. There was a considerable proportion of missing data for each outcome at each time point, varying between 12% and 24%. Consequently, a sensitivity analysis using multiple imputation was conducted and this confirmed the result of the complete case analysis.

The subgroup analysis, although underpowered, suggested that the programme worked better for children who were eligible for FSM compared to their less deprived peers with respect to mid-point literacy ability (Hedges' g=0.22, p=0.03) and the phonics check at post-test (Hedges' g=0.16, p=0.06). There was no statistical evidence that FSM pupils in the intervention group scored better than their peers in the control group on the Woodcock literacy test administered at post-test. Similarly, there was insufficient evidence to suggest that the programme worked better for children with lower baseline attainment (as measured by receptive vocabulary). However, and as noted previously, since the pre-test measure was vocabulary rather than early literacy, it is not possible to know whether children with lower levels of literacy or phonological skills (rather than lower levels of vocabulary) might have benefitted more from the programme.

This is only the third RCT of Success for All to be conducted and it is the first independent trial of the programme in England. The only other trials were based in the US, were not independent and reported a positive effect of the programme, achieving effect sizes in the region of 0.2-0.3. The current trial has been unable to replicate these effects in an English context. Other less robust studies have reported mixed effects of the programme.

The findings from the process evaluation aimed to explore the successes and challenges associated with fidelity and implementation and the findings suggest that fidelity and quality of implementation may be a relevant contributing factor to the null effects. The majority of teachers who took part in the process evaluation reported finding the programme prescriptive and some reported that they adapted the content and delivery. Similarly, Success for All rated the quality of implementation of 15 intervention schools and only one reached the highest 'refined delivery' status after two years, with six schools being classified as 'routine/refined', five as 'routine', and three as 'mechanical/routine'. No schools were rated as 'mechanical' (the poorest level of implementation) by the end of the two-year period. This suggests that as far as quality of implementation is concerned, there was still some way for schools to go and as such, implementation may not have been as successful as anticipated, which may have contributed to attenuating programme effects.

Limitations

It is important to note the potential limitations of the current trial. First, there was attrition of seven schools from the intervention group at the start of the trial. In addition, a further five schools did not fully implement the programme and were presumed by Success for All to have withdrawn from the programme by the end of the two-year delivery period. This means that of the 27 intervention schools, only 15 fully implemented the programme. Although the majority of the withdrawn schools were still included in the intention to treat analysis, this lack of implementation fidelity poses a considerable risk to the observed impact of the programme. An on-treatment analysis suggested that there was some evidence of a small positive effect in favour of the intervention group on the Woodcock literacy test and phonics check at post-test. A further limitation of the trial is the moderate proportion of missing data at the first and second post-test, ranging between 12% and 24%. Consequently, multiple imputation was performed which confirmed the results of the complete case analysis. Finally, the actual analysis was somewhat underpowered due to a smaller than anticipated correlation between the pre- and post-test measures of literacy and this decreased the power of the trial to detect effects smaller than 0.3. The missing data in particular does represent a threat to the generalisability of the findings. An additional limitation was that baseline literacy was not measured as well as baseline vocabulary.

Future research and publications

Future research should focus on efficacy for children from disadvantaged backgrounds as well as ways to enhance implementation fidelity. The programme is clearly enjoyed by those schools who embrace the programme; however, the drop-out post-randomisation (n=7) and the fact that of the 20 remaining schools only 15 implemented it with a reasonable degree of fidelity speaks to the commitment and support required by schools to undertake the programme. The results of this trial will be jointly submitted for publication by the evaluation and project delivery team.

References

Allen, G. (2011). Early Intervention: Smart Investment, Massive Savings. London: Cabinet Office.

Borman, G.D., Slavin, R.E., Cheung, A.C.K., Chamberlain, A.M., Madden, N.A. and Chambers, B., 2007. Final reading outcomes of the national randomized field trial of Success for All. *American Education Research Journal,* 44: 3, 701–731.

Brooks, G., 2007. What works with children with literacy difficulties? The effectiveness of interventions. Slough: National Foundation for Educational Research.

Cowen, J.E. 2003. A Balanced Approach to Beginning Reading Instruction: A Synthesis of Six Major U.S. Research Studies. Newark: International Reading Association.

Dunn, L., Dunn, L., Whetton, C. and Burley, J., 1997. *British Picture Vocabulary Scale*. 2nd ed. London: GL Assessment.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A. and Fletcher, J. M., 1996. Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, 88, 3–17.

Harris, A., Hopkins, D., Youngman, M. and Wordsworth, J., 2001. The implementation and impact of Success for All in English schools. In R.E. Slavin (ed.), *Success for All: Research and reform in elementary education.* Mahwah NJ: Erlbaum.

Juel, C., 1988. Learning to Read and Write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80: 4, 437–447.

National Institute of Child Health and Human Development, 2000. Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Rose, J., 2006. *Independent Review of the Teaching of Early Reading.* London: Department for Education and Skills.

Rowe, K., 2005. *National Inquiry into the Teaching of Literacy (Australia), "Teaching Reading" (2005).* Available at: **http://research.acer.edu.au/tll_misc/5**

Quint, J., Zhu, P., Balu, R., Rappaport, S., and DeLaurentis, M. (2015). Scaling up the Success for All model of school reform. Final report from the Investing in Innovation (i3) Evaluation. New York: MDRC.

Slavin, R. E., Lake, C., Cheung, A. and Davis, S., 2009. Effective reading programs for the elementary grades. *Review of Educational Research*, 79: 4, 1391–1465.

Slavin, R., Wordsworth, J. and Jones-Hill, M., 2005. Success for All in England: Implementation and outcomes of a comprehensive literacy reform for primary schools. Baltimore, MD: Success for All Foundation.

Slavin, R. E., Lake, C., Davis, S., and Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, 6, 1–26.

Torgerson, C. J., 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. Department of Education and Skills Research Report No. RR711. Sheffield: University of Sheffield.

Tracey, L., Chambers, B., Slavin, R. E., Hanley, P. and Cheung, A. (2014). Success for All in England: Results from the Third Year of a National Evaluation. *SAGE Open,* July-September, 1–10.

Tymms, P. and Merrell, C., 2001. *Success for All Quantitative Report: Second Year*. Durham: CEM, Durham University.

Wilson, S. and Lipsey, M., 2007. School-Based Interventions for Aggressive and Disruptive Behavior: Update of a Meta-Analysis. *American Journal of Preventive Medicine*, 33: 2S, 130–143.

Woodcock, R.W., 2011. *Woodcock Reading Mastery Tests*. 3rd ed. Oxford: Pearson.

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[1]					
5	Well conducted experimental design with appropriate analysis	MDES < 0.2	0-10%			Adjustment for Balance [0]		
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%	3		Adjustment for threats to internal validity [0]		3
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	41-50%					
0	No comparator	MDES > 0.6	over 50%					

- **Initial padlock score:** lowest of the three ratings for design, power and attrition = 3 padlocks
- **Reason for adjustment for balance** (if made): N/A
- **Reason for adjustment for threats to validity** (if made): N/A
- **Final padlock score:** initial score adjusted for balance and internal validity = 3 padlocks

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

Appendix C: Consent and information forms

THE UNIVERSITY of York

INSTITUTE FOR EFFECTIVE EDUCATION, THE UNIVERSITY OF YORK

Success for All – Primary School Evaluation

Information Sheet for Parents/Guardians

Dear Parent/Guardian,

We would like to request your permission for your child to take part in an educational research study to evaluate an innovative literacy programme. The following information explains why the research is being done and what it would involve for your child.

What is the Institute for Effective Education?

The Institute for Effective Education (IEE) is part of the University of York. It aims to find out what works in teaching and learning and why, and then use the evidence to improve education.

What is the purpose of this study?

This study is being done to see if using a whole-school literacy programme (Success for All), drawn from research on effective practices in the teaching of reading, helps pupils improve their reading ability.

Why is my child's school participating?

We will conduct this study in 50 primary schools in England. The headteacher of your child's school has agreed to participate in this study.

Does my child have to take part?

You may choose not to permit your child to participate in the testing, but they will participate in the literacy programme, as this will be part of the taught curriculum throughout the school. Your child's test scores will be confidential and will not count towards your child's levels. If you do not want your child to participate, please complete and sign the attached opt-out form by 3:00 pm on (one week later). A pupil's right to withdraw from the testing will be respected.

What will happen in the study?

In September 2013, teachers will administer short literacy tests to all of the Reception pupils in the participating schools. These tests will be followed up with further tests in June 2014 and June 2015. Schools will be randomly assigned to either use the Success for All programme starting in September 2013 and ending in June 2015 or continue their regular literacy teaching.

What will your child's participation be?

Your child will take a short literacy test in September 2013, in June 2014, and at the end of the study in June 2015.

What should you tell your child about the study?

It would be helpful if you could tell your child that the research study is trying to find out what teaching methods help improve children's reading ability.

What are the disadvantages and risks of taking part?

There are no known disadvantages or risks in participating in this study. The tests will be scored at the IEE and will not be shared with anyone. Teachers will continue to teach to the usual literacy objectives throughout the research period.

What are the possible benefits of taking part?

By participating in this study your child will experience a new strategy for learning literacy. Success for All has strong evidence of effectiveness and has been successful in many schools. The information gained from this study may influence how your child will learn in the future.

What happens when the research stops?

IEE researchers will analyse the literacy scores to determine the overall effectiveness of the Success for All programme. Scores for individual pupils and classes will be kept confidential.

When the research is over, the school will receive a report that will show if using Success for All has made a difference to pupils' reading outcomes. The school can then decide whether and how best to use Success for All in the future.

Will my child's information be kept confidential?

Yes. Pupils' names will be replaced with code numbers. No individual pupil's data will appear in any report. Only the research team will not have access to pupils' or parents' names.

What if there is a problem?

If you have a concern or question about your child's participation in this study, please contact Bette Chambers (e-mail: bette.chambers**@york.ac.uk**) Tel: (01904 328153) or Emma Marsden, the head of the Department of Education Ethics Committee (email: misc519@york.ac.uk) about the study.

THE UNIVERSITY of York

INSTITUTE FOR EFFECTIVE EDUCATION

Parent/Guardian opt-out form

If you** do not **permit your child to participate in the study, please complete this form and return it to your child's teacher by 3:00 pm on (one week after receipt).

I **do not** wish my child to take part in the research project.

Pupil's name: .....

Teacher's Name: .....

Parent's/Guardian's name: .....

Parent's/Guardian's signature: .....

Date………

31 May, 2013

SFA Evaluation 2013-2015

Dear colleague,

Thank you for expressing an interest in your school taking part in the 2013-2015 evaluation of the SFA programme, funded by the Education Endowment Foundation. Before we make the final selection of project schools, we would like to give you more information about the programme and the evaluation.

Success for All

Success for All (SFA) is a comprehensive literacy programme for primary pupils, based on research on what works in the teaching of literacy. It has considerable evidence of effectiveness, with studies in many countries, including here in the UK, indicating that it improves children's reading.

In Reception and Year 1, the programme emphasises the teaching of phonemic awareness, systematic phonics, oral language, vocabulary, and comprehension. Pupils work in pairs to help each other learn reading skills, and including letter sounds, sound blending and vocabulary. From Year 2, pupils work in four-member teams reading high quality real books (fiction, non-fiction, poetry, and play scripts) fostering a lifelong love of reading and improving their fluency and reading comprehension. They help each other learn word analysis skills and metacognitive skills such as clarification, questioning, and prediction to comprehend texts from a range of genres. SFA follows the same principles as outlined in Letters and Sounds.

You can learn more about SFA at www.successforall.org.uk.

The Success for All study

This study aims to evaluate the impact of SFA. Participating schools will be randomly assigned either as an "intervention" school or as a "control school". If your school is selected to implement SFA, staff from the SFA-UK registered charity will work with the whole school to implement the SFA curriculum and all training and resources will be provided free of charge.

If your school is selected to take part in the evaluation and is allocated to be a "control" school, SFA-UK will provide the school with £2,000 per year for your participation. Lessons will continue to be taught as normal, and pupils' progress will be assessed by our testers at the end of each year.

Institute for Effective Education Berrick Saul Building University of York YO10 5DD
Tel: 01904 328166 Fax: 01904 328156 Email: iee@york.ac.uk www.york.ac.uk/iee

The evaluation will last two years in total and during this time we will follow Reception pupils (both "intervention" and "control") as they progress through to the end of Year 1. This will involve a one-to-one assessment with each pupil in September/October 2013, and then follow-up assessments at the end of 2014 and 2015.

The enclosed agreement contains more detail. If you are happy to participate in the evaluation on the basis outlined, please sign the enclosed agreement form (a copy of which has also been emailed to you) and return to me by FAX to 01904 328156 or by post to the address below by **Friday 7 June**.

If you do not wish to take the 50% chance of not being assigned to implement SFA, you may contact SFA-UK about purchasing the programme.

With best wishes

Bette Chambers
Director,
Institute for Effective Education

Institute for Effective Education Berrick Saul Building University of York YO10 5DD
Tel: 01904 328166 Fax: 01904 328156 Email: iee@york.ac.uk www.york.ac.uk/iee

Success for All Evaluation 2013-15

This study aims to research the impact of Success for All (SFA), a lively, language rich, literacy programme for primary pupils, based on research on what works in the teaching of literacy. SFA is a complete literacy programme serving pupils from Reception to Year 6.

Participating schools will be randomly assigned either as an "intervention" school or as a "control" school, with a 50% chance of being in either group. Schools selected to implement SFA will be provided with the SFA curriculum and all training and resources free of charge. Control schools will be provided with £2,000 per year for allowing researchers to assess pupils' reading ability.

The evaluation will last two years in total and during this time the study will follow pupils from Reception through to the end of Year 1 (both "intervention" and "control"). This will involve a one-to-one assessment with each pupil in September/October 2013, and then follow-up assessments at the end of the 2014 and 2015 school years. Data from the Year 1 Phonics test will also be analysed. The Year 6 SATs results for individual pupils will also be studied to see how the programme may have affected older pupils in the school.

This trial will provide valuable information to educators and policy makers about effective literacy practices in the UK. All results will remain confidential, with no school being identified by name.

Please sign and return a copy of this form if you agree to take part in the Success for All study and accept the eligibility terms and conditions for receiving the resources and training.

Requirements for schools

- If randomly selected as one of the experimental schools that will be using SFA, teachers will attend training days and work with trainers throughout the project to ensure that they are implementing all aspects of the programme.
- Teachers and school staff will facilitate the visits of IEE staff to conduct individual assessments in September/October 2013, summer 2014 and summer 2015.
- Teachers and schools will make reasonable efforts to accommodate observational visits conducted by researchers.
- Schools will provide the IEE with pupils' scores from the Year 1 Phonics test in 2015.
- Teachers and pupils will complete questionnaires/surveys about aspects of the programme.
- If the school has to withdraw from the study for operational or other unavoidable reasons, wherever possible it will still provide test data for the evaluation.
- Teachers will, at the earliest opportunity, notify IEE if there are any support or operational issues preventing the effective use of the programme.
- Teachers will provide valid email addresses and telephone contact numbers to the researchers and trainers and agree to check communications regularly during the period of the study.

Institute for Effective Education Berrick Saul Building University of York YO10 5DD
Tel: 01904 328166 Fax: 01904 328156 Email: iee@york.ac.uk www.york.ac.uk/iee

- The school should not already be participating in another literacy research project or programme evaluation.

Commitments of the IEE

- All data provided by schools will be stored in accordance with the Data Protection Act (1998).
- All results will be anonymised so that no schools or individuals will be identifiable in the report or dissemination of results. Confidentiality will be maintained and no one outside the evaluation team will have access to the database.
- Training, resources and ongoing support will be provided by Success for All – UK to all implementation schools free of charge.
- Control schools will receive £2,000 per year for their participation.

Headteacher agreement

I agree for my school _____ to take part in the SFA study and I accept the eligibility terms and conditions.

Signature of Head Teacher: _____

Name of Head Teacher:_____

School name: _____

Date: ___/___/_____

Signature of Researcher: _____

Name of Researcher:_____

Date: ___/___/_____

Institute for Effective Education Berrick Saul Building University of York YO10 5DD
Tel: 01904 328166 Fax: 01904 328156 Email: iee@york.ac.uk www.york.ac.uk/iee

Appendix D: Fidelity rating form

Introduction

"...effective implementation is associated with better outcomes"

Durlak & DuPre (2008)

This document is a guide for School Leaders and Facilitators so that they have a clear understanding of the level of Implementation Fidelity in their school which is **critical to the success of the programme** and the outcomes for children. It is also an aide memoire to support the continuous development of the programme in school at all levels. Some of the evidence for these evaluations should be available through the completion of your SFA Leadership Snapshot (available through your lead consultant)

Implementation is classified into 3 stages, these are:

1- Mechanical
2- Routine
3- Refined

Schools that see the biggest impact from the use of SFA are those that can progress from Mechanical to Refined as quickly as possible.

It is important to continually evaluate the school's level of implementation addressing the Mechanical, Routine and Refined areas of Fidelity in a systematic way. It is important that these are viewed as strategic building building blocks that will give the school the success in outcomes
it requires.

There is the likelihood that whilst completing the evaluation of Mechanical aspects of the programme you will observe that the school is beginning to see the evolution of aspects of higher levels of Implementation. This is to be expected as each school will progress at a different rate. You may find that your school is meeting some Routine and Refined elements whilst you are working towards an overall rating of Mechanical.

"If programmes are not implemented with fidelity, potential returns are threatened...which is why I have established a list of the best evidence-based programmes... Fidelity is key to the success of Early Intervention programmes."

Allen (2011)

Snapshot Ratings

For all objectives, the rating is either:

- **In Place (I)**
- **Not In Place (N)**

When used in conjunction with the SFA Leadership Log (found in the Leadership Snapshot), for some Schoolwide Structures objectives, an (S) rating may be used to indicate that the school has met start-up expectations.

For Instructional Processes (IP) and Pupil Engagement (PE) objectives, the final rating is summative.

After rating individual classrooms, the final rating for each component (Curiosity Corner-CC, Kinder Corner-KC, Roots, Wings or Quest) is based on how many classrooms had objectives in place:

P = **Power schoolwide** (Objective is verified for 95% of teachers.)
M = **Mastery** (Objective is verified for 80% of teachers.)
S = **Significant use** (Objective is verified for 40% of teachers.)
L = **Learning** (Staff members are working toward verification of this objective.)

Ratings may be based on observation, discussion, classroom samples, and other artifacts. Ratings are verified by observation or artifacts, such as team score sheets, facilitator observation records, videos, audio records, transcripts of instruction, or teacher records of student responses.

A school may find that its evaluation profile is different across components for various reasons.
Eg. A school may be **P** in some areas but **L** in others.

This self evaluation will help the Facilitators identify with their Consultants and Senior leaders, priorities for development. (Refer to the document **'Leading for Success- Teaching and Learning: Hints and Tips for Improving Implementation Fidelity'** for more detail on how to find evidence for each area). This information can be used to identify areas of strength and as a tool to identify and support areas of development and plan professional development for individuals and teams to help the school towards a Power level of implementation.

Mechanical: The First Stage of Implementation

In this Table **T** refers to **Teacher** and the number of columns can be adjusted to show the number of teachers delivering the component. One of these sheets should be completed for each component used in the school (Curiosity Corner, Kinder Corner, Roots, Wings or Quest)*

*For more information on any of the programs not currently in use in your school, please speak to your lead consultant.

IP = Instructional Process (Teaching) **PE = Pupil Engagement** (Learning)			**IP = In place N = Not in place**								**Summary P/M/S/L**
			T1	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	
Mechanical	**IP**	Teachers use the basic lesson structure and objectives. Teachers use available media regularly and effectively. **(IP-1)**									
	PE	Pupils are familiar with routines. **(PE-1)**									

Routine: The Second Level of Implementation

As you would expect, there are more detailed requirements that need to be met to achieve a 'Routine' Evaluation, some of which you will already have seen evidence of whilst reviewing the Mechanical implementation. It is in this stage that schools see the most rapid improvement in progress and attainment.

IP = Instructional Process (Teaching) **PE = Pupil Engagement** (Learning)			**IP = In place N = Not in place**								**Summary P/M/S/L**
			T1	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	
Routine	**IP**	Teachers provide sentence stems when asking questions and model appropriate responses for children to structure their answers in full sentences. **(IP-3)**									
		Teachers restate and elaborate pupil responses to promote vocabulary mastery at a high standard of oral expression. **(IP-5)**									
		During partner reading teachers listen to, observe and collect information (running records and/or informal fluency evaluations) on a focus group of children and / or team each lesson. **(IP-7)**									
		Teachers refer to the types of questions (Copy Cat, Text Detective, Judge and Jury) and model strategies to answer each question type during Shared treasure in Roots and Treasure Hunt questions in Wings **(IP-9)**									
		Teachers calculate team scores and celebrate team success in every cycle. **(IP-13)**									
		Read and Respond forms are collected weekly, and return is celebrated. Return rate is 80% or better. **(IP-15)**									
	PE	Pupil talk equals or exceeds teacher talk. (Each pupil should be engaged in partner/team discussion as a speaker or active listener during half of class time.) **(PE-3)**									
		Pupils are engaged during team/partner practice and labs. If needed, strategies such as talking chips or role cards are in use. **(PE-4)**									
		Pupils value team scores and work daily to ensure that team members are prepared to successfully report for the team during Random Reporter and to succeed on tests. **(PE-8)**									

Refined: This is the highest level of Implementation Fidelity

This is the most demanding stage which leads to long term sustainability of SFA and increased outcomes.

IP = Instructional Process (Teaching) **PE = Pupil Engagement** (Learning)			**IP = In place N = Not in place**							**Summary P/M/S/L**
			T1	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	
Refined	**IP**	Active instruction is appropriately paced and includes modelling and guided practice that is responsive to pupils' understanding of the objective. **(IP-2)**								
		Teachers use Think-Pair-Share, whole-group response, Random Reporter (or similar tools that require every pupil to prepare to respond) frequently and effectively during teacher presentation. **(IP-4)**								
		Teachers provide time for partner and team talk and lab activities to allow mastery of learning objectives by all pupils. **(IP-6)**								
		Teachers facilitate partner discussion and pupil interaction by circulating, questioning, redirecting, and challenging pupils to increase the depth of discussion and ensure individual progress. **(IP-8)**								
		Following Team Talk or other team study discussion, teachers conduct a class discussion in which pupils are randomly selected to report for their teams; team points are awarded. **(IP-10)**								
		During class discussion, teachers effectively summarise, address misconceptions or inaccuracies, and extend thinking through thoughtful questioning. **(IP-11)**								
		During class discussion, teachers ask pupils to share both successful and unsuccessful use of strategies, such as clarifying, questioning, predicting, summarising, and graphic organisers. **(IP-12)**								
		Teachers use team scores to help pupils set goals for improvement, and pupils receive points for meeting goals. **(IP-13)**								
		Teachers use team scores to help pupils set targets for improvement and pupils receive points for meeting targets. **(IP-14)**								
		Teachers conduct Class Council meetings weekly. The atmosphere is open, and relevant class issues are addressed effectively. **(IP-16)**								
		Teachers facilitate the use of emotion-control and conflict-resolution strategies throughout the day. **(IP-17)**								

IP = Instructional Process (Teaching) **PE = Pupil Engagement** (Learning)			**IP = In place N = Not in place**							**Summary P/M/S/L**
			T1	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	
Refined	**PE**	Pupils speak in full, elaborate sentences when responding to teacher questions. **(PE-2)**								
		Partners assist each other effectively with difficult words and use re-tell every day during partner reading. **(PE-5)**								
		Pupils use rubrics to meet expectations (e.g., fluency, writing, vocabulary, strategy use, comprehension). **(PE-6)**								
		Teams are engaged in highly challenging discussions, in which pupils explain and offer evidence from the text to support their answers, or, for writing, pupils offer thoughtful responses during the revision process. **(PE-7)**								
		Pupils use strategy cards to assist one another during reading and discussion & revision guides to offer helpful feedback during the writing process. **(PE-9)**								
		Pupils know their reading levels and can articulate what they need to do to increase their reading achievement, or, for writing, pupils know their reading and writing strengths and what they need to do to improve their writing. **(PE-10)**								
		Pupils use decision making skills to solve problems that arise. **(PE-11)**								
		Pupils can identify the intensity of their feelings and use self-control strategies when needed. **(PE-12)**								

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

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