



E
E
F
ducation
ndowment
oundation

Nuffield Early Language Intervention

Evaluation report and executive summary

February 2016

Independent evaluator:

Luke Sibieta (Institute of Fiscal Studies)



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 the Institute for Fiscal Studies. The lead evaluator was Luke Sibieta.

Contact details:

The Institute for Fiscal Studies
3rd Floor
7 Ridgmount Street
London
WC1E 7AE

p: 020 7291 4800

e: mailbox@ifs.org.uk

Contents

About the evaluator.....	2
Contents.....	3
Executive summary.....	4
Introduction	6
Methods	11
Impact evaluation	22
Process evaluation.....	36
Conclusion.....	43
References.....	46
Appendix A: Additional tables	48
Appendix B: Subsample balance analysis.....	49
Appendix C: Alternative methodologies	54
Appendix D: School and parent consent letters.....	55
Appendix E: Topic guide for programme coordinators	66
Appendix F: Topic guide for teaching assistants.....	72
Appendix G: Cost rating	78
Appendix H: Security classification of trial findings.....	79

Executive summary

The project

The Nuffield Early Language Intervention is designed to improve the spoken language ability of children during the transition from nursery to primary school. It is targeted at children with relatively poor spoken language skills. Three sessions per week are delivered to groups of two to four children starting in the final term of nursery and continuing in the first two terms of reception in primary school. Children in primary school also attend an additional two 15-minute individual sessions per week. All sessions focus on listening, narrative and vocabulary skills. Work on phonological awareness is introduced in the final ten weeks.

The intervention was developed by researchers from the University of York with funding from the Nuffield Foundation. The communications charity I CAN was enlisted to train teaching assistants and nursery staff to deliver the programme. This report evaluates the I CAN-led model for the 30-week programme described above and also a shorter 20-week version delivered only in reception year.

The impact of these two programmes on the language skills of 350 children in 34 schools was tested using a randomised controlled trial design. Schools with attached nursery schools or nursery classes in Yorkshire and the South East were recruited to the trial in 2013. Children identified as having relatively low language skills were randomly allocated to the 30-week programme, the 20-week programme or standard provision. The qualitative fieldwork carried out as part of the project involved interviews with a total of 12 staff in 8 of the 34 participating schools.

Key conclusions

1. The Nuffield Early Language Intervention had a positive impact on the language skills of children in the trial. This is true for both the more expensive, 30-week version, starting in nursery, and the 20-week version, delivered only in school.
2. Children receiving the 30-week version experienced the equivalent of about four months of additional progress, compared with about 2 months additional progress for the 20-week version. Both results are unlikely to have occurred by chance, though results for the 30-week version are more secure.
3. The evaluation did not provide reliable evidence that either version of the programme had a positive impact on children's word-level literacy skills.
4. Teaching assistants delivering the programme reported that they found it difficult to devote enough time to it, and that support from senior staff was required to protect the programme time.
5. Staff in participating schools reported that the programme had a positive impact on children's language skills and confidence. They thought that the factors which contributed to this included the small-group format, the activities covered, and the focus on narrative and vocabulary work.

Security rating

Security rating awarded as part of the EEF peer review process

Findings from this trial have moderate to high security. The trial was set up as a randomised controlled trial that aimed to compare the progress of children who received the interventions with that of similar children who did not. Randomisation was at the child level within nurseries/schools. The trial is classified as an efficacy trial because it tested whether the intervention can work under ideal or developer-led conditions, but did not seek to demonstrate that the approach would work in all types of schools. The trial was large and well-conducted. The 'padlock' security rating is 4 rather than 5 because 11% of randomised pupils did not complete all the tests at the end of the project.

Findings

Children receiving the 30-week version experienced the equivalent of about four months of additional progress. This effect is unlikely to have occurred by chance. For the 20-week version the figure was smaller, equivalent to about two months, and slightly more likely to have occurred by chance.

The evaluation also measured the impact of the interventions on children six months after it had ended. The results suggest that the impact may have actually increased over time (for both versions). However, it is not possible to say this with confidence because not enough is known about what other activities the children were involved in during the six-month follow up period.

On average, children who have better language skills also have better literacy skills, so it might be expected that if the programmes improved language skills then they would also improve children's literacy skills. However, this evaluation provided no reliable evidence that either version of the programme had a positive impact on children's word-level literacy skills in the short term.

Staff reported improvements in the spoken language skills, conversational ability, narrative skills, and vocabulary of the participating children as well as improved listening skills and better general language development. Participating children were also perceived to be more confident, outgoing and conversational after taking part. Staff thought that the factors which contributed to the effectiveness of the programme were: the small group format and regular sessions; the fun and engaging nature of the content; the focus on appropriate skills; and the repetition of knowledge across sessions. The process evaluation found that effective delivery of the programmes required staff to have adequate delivery and preparation time, and the school to have a separate space where group and individual sessions could be delivered. Teaching assistants (TAs) reported finding it hard to deal with the additional workload, and so schools might need to consider ways to ensure that provision of the programmes does not adversely impact other services provided by teaching assistants.

How much does it cost?

The programme is relatively cheap to buy but requires significant delivery time from TAs. The cost of providing one TA with training and materials to deliver the 30-week intervention is just under £2,500; for the 20-week intervention it is £1,400. Each TA could then deliver the intervention repeatedly. In terms of staff time, for the 20-week intervention the requirement is 4.5 hours per week for 20 weeks (90 hours) per group of four children. For the 30-week intervention there are an additional ten weeks in nursery requiring two hours per week per group of four children making a total requirement of 110 hours. These time requirements include some preparation time, although the process evaluation suggested additional preparation time is needed in practice.

Table 1: Summary of impact on language skills

Group	Effect size (95% confidence interval)	Estimated months' progress	Security rating	EEF cost rating
30-week vs. untreated control	0.27 (0.07; 0.46)	4 months		£
20-week vs. untreated control	0.16 (-0.02; 0.34)	2 months		£

Note: See notes to Table 11 for more details on how the impact estimates were calculated. See the 'cost' section in the 'Impact evaluation' chapter for more detail on the EEF cost rating.

Introduction

Intervention

The aim of this evaluation was to test whether the provision of the Nuffield Early Language Intervention improved the language and literacy skills of children with relatively low language abilities. Both a 30-week and 20-week treatment were administered as part of the intervention to identify whether a longer programme that starts in nursery and continues during reception year in primary school is more effective than a shorter programme that starts when children enter primary school.

The intervention is designed to be delivered by teaching assistants. It involves staff training, a detailed set of lesson plans, and materials for three ten-week blocks of teaching. Two versions of the intervention were included in the trial. Pupils in one treatment group received a 30-week programme starting in the final term of nursery and continuing for the first two terms of reception year in primary school. Pupils in the second treatment group received a 20-week programme that ran during the first two terms of primary school.

In nursery, the intervention was delivered to groups of two to four children during three 20-minute sessions per week. In reception, these group sessions with two to four children were extended to 30 minutes and complemented with two 15-minute individual sessions per week. Figure 1 shows the overall structure of intervention delivery.

Figure 1: Structure of programme delivery in schools

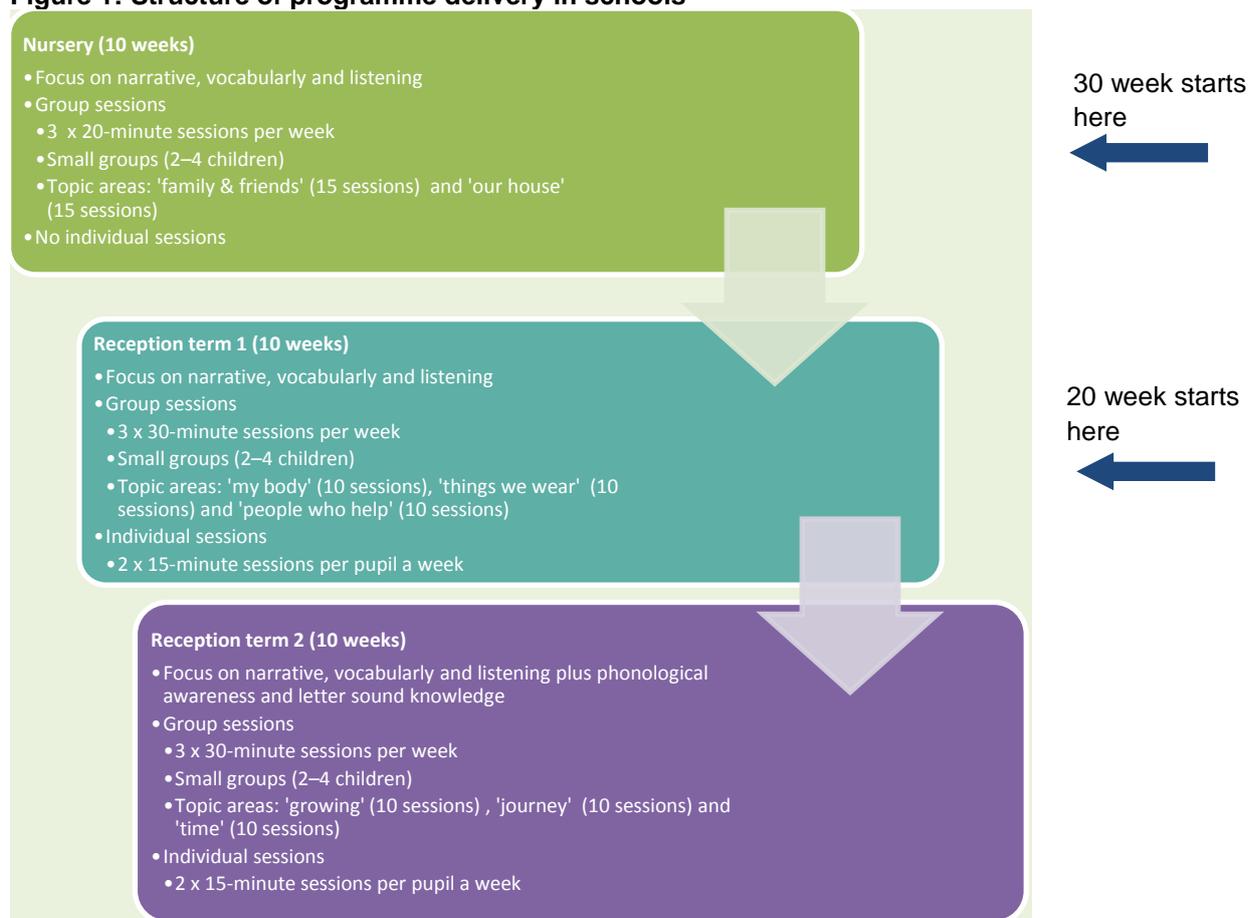
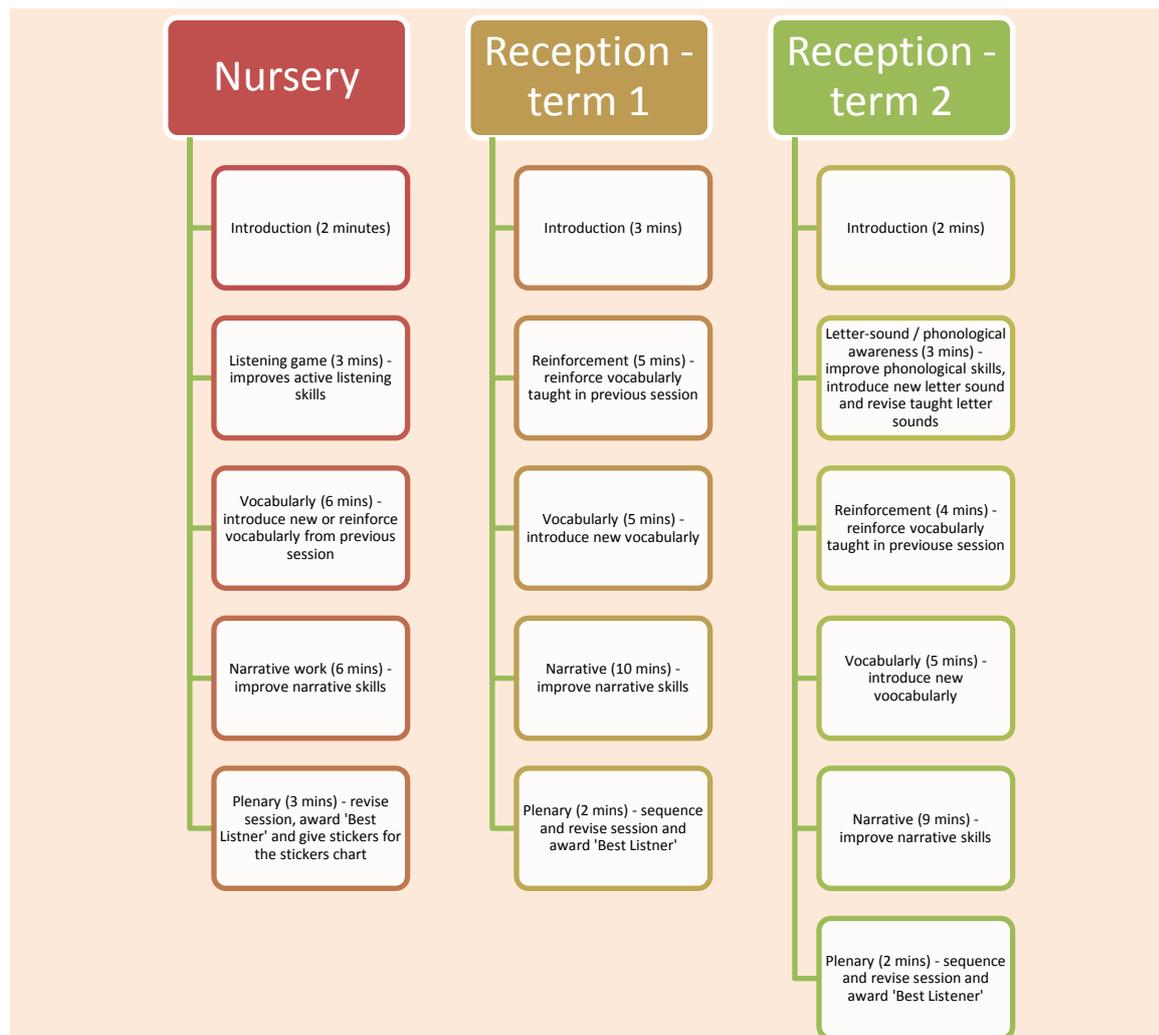


Figure 2 outlines the key components that were delivered in each session in order to develop and consolidate vocabulary, listening and narrative skills. An added focus on phonological awareness and letter sound knowledge was introduced in the second term of reception. Within this repetitive framework, multi-sensory techniques were used (for example flashcards associated with words) to help make learning enjoyable and to encourage active interaction between the children themselves and with the TA. Rewarding children was an integral feature of each session; this involved targeted verbal praise and more formal praise in the form of a Best Listener Award given to the child that has listened well in the class, and stickers given to the rest of their class for their effort.

Figure 2: Components of the group sessions



The staff trained to deliver the intervention were mainly TAs, but also included some senior school staff. We do not know whether these senior staff went on to deliver the intervention or just attended the training. Training consisted of two events that were arranged by the programme team. There was a one-day event for the nursery component (required for the 30-week intervention) and, later in the year, a two-day event for the reception component (required for both the 20- and 30-week interventions). Apart from covering intervention delivery, these events also explained the purpose of the programme and the specific guidelines and restrictions governing the trial—such as the importance of delivering the programme according to the manual and not using the programme with non-intervention children. Children who attended a nursery that was attached to a participating primary school were selected to take part in the trial if they demonstrated low language skills as

measured by the CELF expressive vocabulary and sentence structure tests.¹ Within each nursery, children were then randomly allocated to either a 30-week intervention, a 20-week intervention or the control group (the randomisation is described in more detail in the ‘Methodology’ section below). Children in the control group received no additional language support during the trial beyond that normally received in a business-as-usual scenario, but their language and word-level literacy skills were monitored in the same way as the two treatment groups. Schools in the trial were given the opportunity to deliver the Reading and Language Intervention (RALI) to pupils from the control group after the end of the trial. RALI has already been shown to improve word-level literacy skills.

Background evidence

It is well known that children with oral language difficulties often struggle to read well. In particular, poor language skills are often associated with poor reading comprehension, which is fundamental to the development of functional literacy (Clarke *et al.*, 2010). Given the importance of basic skills such as literacy for an individual’s subsequent educational attainment and labour market success, there is substantial interest in understanding how to improve children’s literacy skills.

The existing academic literature shows a strong relationship between the development of language and literacy skills (Whitehurst and Lonigan, 1998; Scarborough, 2009). For example, understanding words when spoken makes it easier for children to read the same words in text. Early literacy and language skills are strongly correlated with one another, but also predict future reading ability. The relationship between non-verbal skills—such as visual memory and motor skills—and later literacy is, by contrast, much weaker (Scarborough, 2009). There is also consensus that the process of learning to read begins early in the pre-school period. Once at school, deficits in literacy skills can be difficult to address, which has led to the conclusion that those with poor language or literacy skills should receive help as early as possible (Whitehurst and Lonigan, 1998; Scarborough, 2009; Hulme and Snowling, 2013).

Earlier randomised controlled trials have found evidence that oral language interventions can achieve improvements in skills that are strongly related to reading comprehension ability (Bowyer-Crane *et al.*, 2008; Snowling and Hulme, 2011). To further examine these results, the Nuffield Foundation funded a team from the University of York to develop a programme based on the oral language interventions used in these earlier studies.

An initial version of the programme consisted of a 30-week intervention delivered by teaching assistants that began during nursery and continued into reception. It was delivered in small groups complemented with individual sessions and focused on spoken language skills such as listening comprehension and vocabulary. A trial of this intervention found statistically significant and positive impacts on oral language and spoken narrative skills (Fricke *et al.*, 2013).² In particular, the study found:

- very large and statistically significant positive effects on an overall measure of language skills (reported effect sizes of about 0.80–0.83);
- large and statistically significant positive impacts on narrative skills (reported effect sizes of 0.30–0.39) and phoneme awareness (reported effect size of 0.49);
- statistically insignificant effects on overall literacy skills (reported effect sizes of 0.14–0.31; and
- evidence of a large and statistically significant effect on reading comprehension (reported effect size of 0.97).

¹ See <http://www.pearsonclinical.com/language/products/100000316/celf-preschool-2-celf-preschool-2.html#tab-details>

² In its development stage, the Nuffield Early Language Intervention was called Nuffield Language4Reading.

This pilot study was about half the size of the present study (180 pupils across 15 nursery schools), but was very similar in design and used similar screening tests.

Following this study, the Nuffield Foundation has worked with the communication charity I CAN to market the intervention and deliver training to licensed trainers and TAs. In this evaluation we examine the effects of this current form of the 30-week Nuffield Early Language Intervention. As such, our evaluation represents an efficacy trial of the intervention model as it would be delivered at scale. In addition, we compare this with the effects of a shorter 20-week version of the trial (administered during reception year in primary school only). This shorter version was included in the evaluation as it potentially represents a lower cost, and therefore more scalable, version of the intervention.

Evaluation objectives

The main research questions for this evaluation are as follows:

- 1) What is the impact of receiving 20 or 30 weeks intensive language support at age four or five on children's vocabulary, reading, spelling and comprehension skills one year later?
- 2) How much more effective is it to receive 30 rather than 20 weeks support?
- 3) To what extent do differences in the effects of the 20- and 30-week interventions [relative to the control group] continue or fade out once the programme has finished?

In addition to these questions the process evaluation explored participants' views and experiences of the delivery, impact and perceived value of the programme.

The evaluation protocol was published on the EEF website:

<https://educationendowmentfoundation.org.uk/projects/language-for-learning/>

Ethical review

Ethical approval was obtained from the UCL ethics board for the intervention and impact evaluation undertaken by IFS researchers. IFS researchers were also required to adhere to the Economic and Social Research Council's Ethics Framework, the Social Research Association's Ethical Guidelines, as well as the IFS Information Security guidelines and the IFS Information Classification and Handling Policy (both of which comply with the international standard for data security, ISO27001).

Schools gave initial permission for all pupils in the final year of the attached nursery to be given the screening test. Pupils were identified as eligible for the intervention if their score was among the lowest 12 scores in school. If a pupil was identified as eligible, their parents/carers were asked for informed opt-in consent for the pupil's participation in the research project. The consent request included information about the nature of the project, what data would be collected, linkage of data to the National Pupil Database, and how the data would be used. Parents were also regularly informed about the progress of the trial, the contact details of the project team, and the option to withdraw their child at any point. (A copy of one of the updates sent to parents and schools is included in Appendix D.)

Participating schools were asked to provide Unique Pupil Numbers for participating pupils to enable pupil test scores to be linked to National Pupil Database (NPD) records. Most schools provided them. A small number refused to provide this information; others failed to respond to repeated requests from the project team for this information. After the evaluation report is published, further efforts will be made to contact those schools that did not respond in order to ensure that as many pupils as possible can be tracked over time using the NPD.

NatCen Social Research obtained ethical approval from their internal ethics board for the process evaluation. Approval was sought regarding communication with schools, the opt-out process, and the fieldwork.

Details of all initial information sheets and consent forms for schools and parents are included in Appendix D.

Evaluation team

Lead researcher from IFS:

Luke Sibieta, Programme Director at IFS

Supported by:

Claire Crawford, Research Fellow at IFS

Elaine Kelly, Senior Research Economist at IFS

Agnes Norris Keiller, Research Associate at IFS

Process evaluation team from NatCen Social Research:

Amy Skipp, Research Director of the Children and Young People team

Mehul Kotecha, Senior Researcher

Project and delivery team

Professor Charles Hulme, University College London

Dr. Silke Fricke, University of Sheffield

Dr. Claudine Bowyer-Crane, University of York

Mandy Grist, I CAN Communication Advisor

Professor Margaret Snowling, University of Oxford

Trial registration

The trial was not registered by the evaluation or project delivery teams.

Methods

Trial design

The overall aims of the trial were to test the impact of the existing 30-week and new 20-week versions of the Nuffield Early Language Intervention on language and literacy skills, and to explore the participants' views on the delivery, impact, and perceived value of the intervention. This was an efficacy trial, designed to test the delivery of the intervention at relative scale, but with support from the developers. The evaluation seeks to estimate the effects of the interventions on an intention-to-treat basis, which means that schools and pupils that dropped out will be included in the analysis wherever possible.

The trial was designed as a randomised controlled trial, with randomisation at the pupil level within nurseries attached to schools. The randomisation was undertaken independently by members of the impact evaluation team who randomly allocated pupils within each nursery to one of three groups—a control group or one of two treatment groups. The methods used are described in the 'Randomisation' section below. Pupils in each treatment group received either a 30- or 20-week version of the intervention.³ A waitlist control group was chosen to address the ethical issues of identifying a group of struggling pupils and then not providing them any additional support. Schools in the trial were given the opportunity to deliver the Reading and Language Intervention (RALI) to pupils from the control group after the end of the trial. RALI is an intervention that has already been shown to improve literacy skills. Schools were offered the choice of training to deliver RALI to control group pupils or a one-off payment equal to the cost of delivering RALI so that they could deliver an alternative intervention of their choice. About half the schools chose the RALI option, and about half opted for the cash payment.

The main reasons for adopting within-school randomisation, as opposed to an across-school randomisation, were the risks created by differential attrition⁴ and the improved statistical power resulting from within-school randomisation. The main downside of within-school randomisation is that spillover effects are more likely resulting from, for example, TAs applying the intervention techniques to pupils in the control group. In such a case, if pupils in the control group are impacted positively the impact estimates would *underestimate* the impact of the treatments. In light of this concern, the training emphasised the importance of maintaining treatment and control conditions, and the process evaluation explicitly asked TAs about whether they had used skills gained from the trial with other pupils. The process evaluation suggests that participants were fully aware of the importance of the experimental conditions and sought to maintain them. We therefore speculate that such spillover effects are likely to be small, but we are unable to rule out the possibility. Another potential spillover effect was the effect of the high workload involved in delivering the trial. If this had a *negative* impact on the control group – because they were receiving less teaching assistant resource than they usually would – then the impact estimates would *overestimate* the impact of the treatments. ,

The process evaluation identified several minor and more significant deviations from the prescribed model. The project team collected data about the delivery model used by schools in the trial and reported that 5 of the 31 schools that completed the intervention showed significant deviations related to both the structure and delivery of the programme. These included a delayed start, failing to deliver three group sessions per week, and the merging of children on the 20- and 30-week versions of the programme into a single group.

³ Pupils receiving the 20-week intervention received no additional support during the summer term in nursery prior to entering reception.

⁴ If we had adopted an across-school randomisation, then the drop-out of whole schools could have led to significant imbalances in the characteristics of treatment and control pupils, particularly as the number of schools involved was relatively small (for example, if a particularly deprived school had dropped out, then the groups could have easily become imbalanced on this characteristic).

Minor deviations tended to relate to the structure of the programme and usually involved TAs increasing or reducing the length of group or individual sessions. Deviations relating to the individual sessions were more common and included providing fewer than the prescribed two per week, reducing the duration of the sessions, or delivering them in a general classroom while the TA was also supervising the class. Deviations in the delivery of the programme included removing key lesson components and changing the resources used. Despite these deviations, the process evaluation found the delivery of the programme structure and session components to be generally in line with the prescribed model (see the 'Process evaluation' section for more detail).

Group and individual sessions were usually delivered during lesson time meaning that pupils were taken out of another class. For some this was a different class each week, for others it was the same class each week. Occasionally, sessions were delivered during unstructured learning time (such as during free-play sessions) in order not to disrupt other lessons. We should therefore interpret the effects of the interventions relative to a business-as-usual scenario where pupils would have spent the time in the classes.

Owing to the high frequency of sessions, it was common for pupils in the treatment groups to miss several sessions over the course of the intervention. The project team asked TAs to record the number of sessions/hours received by each individual pupil. According to these records, pupils, on average, attended just over 80% of the group sessions in both nursery and reception phases, and 56% of individual sessions in reception. It is not clear whether these figures represent actual attendance levels or whether they are the result of a failure to record attendance accurately. These figures, therefore, should be treated as a lower bound on actual attendance levels.

Such partial participation in the treatment would be problematic if our focus was the average treatment effect on the treated (ATT). By ignoring partial programme participation, this parameter would overstate the impact of the treatments that could be expected if the programmes were repeated. To avoid this, all results presented here are estimated on an intention-to-treat (ITT) basis, which means the partial participation does not pose a problem.

Participant selection

The intervention focused on primary schools with attached nurseries. This decision was taken as the 30-week version of the intervention bridged the transition from nursery to reception year in primary school and it was important to ensure that pupils in participating nurseries would progress into participating primary schools. To facilitate the implementation of the programme, schools were only considered if they were relatively close to the project team which was based in Yorkshire and the South East. Schools in disadvantaged areas were targeted in line with the EEF focus on disadvantaged pupils. Schools with high proportions of pupils with English as an Additional Language were not targeted as the intervention was hypothesised to be of less benefit to such pupils.

Using these criteria I CAN managed to recruit 34 schools in total, 4 more than the initial target of 30 schools. This is out of a total of 302 schools approached. The recruitment process began with an initial email and letter, which was then followed up with phone calls in most cases. The major barriers to recruiting schools to participate included finding the 'right person' in the school, schools being unable to commit to the amount of staff time that would be needed without funding, schools needing to focus on existing priorities (i.e. working towards academy status) or already having training/interventions running and needing to maintain focus on them.

Within the schools (with attached nurseries) that agreed to participate, eligible pupils were identified using a screening process. The intervention was focused on improving the spoken language abilities of children with relatively poor language skills. To identify eligible children, a composite measure of language skills was constructed as the average of the CELF expressive vocabulary and sentence

structure tests.⁵ In each nursery, the parents of the children with the 12 lowest composite language scores were invited to give consent for their child to participate in the trial. If the parents of any of these children opted out of the trial, the parents of the next lowest scoring child were approached until the nursery had enlisted an adequate number of children (the lowest number of participating children in one nursery was nine).

Outcomes measures

The primary outcome for this evaluation is a language skills score which is a composite score of four different externally-valid measures of language skill:

- **Renfrew Action Picture Test (APT)** (revised Renfrew, 2010). In this test, the child is asked to describe the actions shown in a set of pictures. Two scores are recorded, one for the level of information they provide (for example nouns and verbs) and one for the grammar they use (such as use of tenses). We use both components.
- **CELF-Preschool 2 UK: Expressive Vocabulary.**⁶ In this test, children are asked to name pictures.
- **Listening Comprehension** (based on the York Assessment of Reading Comprehension test, YARC). Children listen to recordings of four short stories and answer questions about them.⁷

In addition, we define a secondary composite outcome of word-level literacy skills based on the following three measures:

- **YARC: Letter Knowledge.**⁸ This requires children to say the sounds of letters.
- **YARC: Early Word Reading.** This requires children to say the sounds of simple words aloud.
- **Spelling:** children are asked to write a series of simple words.

All these component measures are standardised, age appropriate, and were chosen by the project team to be consistent with the measures used in their Randomised Controlled Trial (Fricke *et al.*, 2013), and the aims of the intervention to improve language and literacy skills. However, we are not aware of any evidence that suggests these specific tests are predictive of future performance in national tests. All tests were administered and scored by research assistants trained by the project team who were blind to the allocation of children to groups.

Tests were conducted at three main intervals during the course of the trial:

- **Pre-test:** this was conducted in April 2013 just before the start of the intervention phase.
- **Post-test:** undertaken between May and July 2014 after the end of the intervention phase.
- **Follow-up test:** these tests were undertaken about six months after the end of the intervention phase between October and December 2014. By this stage some control pupils had already begun to receive the alternative intervention (RALI). This is not ideal if we are interested in the extent to which differences across treatment and control pupils persist after the end of the intervention. In the 'Analysis' section below, we go into more detail about what we can learn from analysis of the follow-up tests.

⁵ Semel, Wiig, and Secord (2003). A thorough description of the test is provided by Paslawski (2005).

⁶ Semel, Wiig, and Secord (2003).

⁷ The listening comprehension test is a measure created by the project developers. The stories and questions are published in the YARC test with a reference as follows: Snowling, M.J., Stothard, SE., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., Nation, K., & Hulme, C. (2009) YARC York Assessment of Reading for Comprehension. Passage Reading. GL Publishers.

⁸ <http://www.gl-education.com/international-products/york-assessment-reading-comprehension-yarc>

Descriptive statistics of the raw pre-test and post-test scores are shown in Table 2. The various raw component scores had very different scales and standard deviations, such that simply adding them directly would not be meaningful. We therefore standardised each of the components to have mean zero and standard deviation one. These were then added together to create composite measures and re-standardised. This means that each component contributed equally to the composite measure. This same process was applied to create both the primary composite score of language skills and the secondary composite score of word-level literacy skills. This composite approach has the advantages that the contribution of each component to the total score is equal and is not dependent on the scales used in each component test. However we also present differences in each component by treatment and control groups to allow readers to make their own judgements about the relative importance of each component.

Table 2: Descriptive statistics of raw outcome measure components

Measure	Pre-test					Post-test				
	mean	min	max	s.d.	N	mean	min	max	s.d.	N
Language skill measures										
CELF expressive vocabulary	10.80	0	28	5.33	394	20.46	3	36	5.90	352
APT information	20.39	0	35.5	6.15	394	27.98	11.5	37.5	4.66	352
APT grammar	13.92	0	31	5.96	394	22.02	7	34	5.01	352
Listening comprehension	1.34	0	8	1.59	391	4.80	0	11	2.65	352
Word-level literacy skill measures										
YARC letter sound knowledge	1.84	0	14	2.76	394	26.87	3	32	5.07	352
YARC early word reading	0.18	0	21	1.39	394	9.27	0	30	6.65	352
Spelling total	2.24	0	40	4.26	389	58.07	0	128	27.75	353

Sample size

The project team aimed to recruit a total of 360 pupils across 30 schools, with pupils spread equally across the two treatment groups and the control group in each school. Our initial power calculations assumed a central scenario where 20% of the variation in post-test was explained by pre-test characteristics (also assuming 80% test power and 5% significance level). In this central scenario a sample size of 360 would be sufficient to detect an effect size of 0.32 SDs. In the event that the pre-test characteristics explained none of the variation in post-test scores, the sample would only be sufficient to detect an effect size of 0.37 SDs. Conversely, were the pre-test characteristics to explain 40% of the post-test outcomes, the minimum detectable impact would fall to 0.29 SDs.

The project team actually managed to recruit four more schools than planned, and a total of 394 nursery pupils were included in the intervention in spring/summer 2013. Three schools subsequently withdrew from the intervention and 27 pupils in participating schools dropped out because they changed schools between nursery and reception. In schools that dropped out, pre and post-test score

data was still collected for pupils. However, if pupils moved school, then we were not always able to collect test score data. Test score data was also missing for a small number of pupils at either pre-test or post-test stages.

In light of this attrition and missing data we imposed a common sample, dropping pupils with any missing outcomes at either of the two test points. This ensures that differences in the estimated impact of the interventions across outcomes are not due to changes in the pupils included in the sample. Although some bias may be introduced as a result, this is likely to be minimal as creating this common sample only involved dropping three cases with some non-missing data.

The final sample of pupils that appear in both nursery and reception in participating schools with non-missing data was 350, ten fewer than originally planned.

Randomisation

The randomisation was undertaken by members of the independent impact evaluation team. The randomisation process allocated pupils within each nursery to two treatment and one control group within each nursery/school and explicitly sought to minimise differences across groups in terms of age, gender and the screening test score. However, the actual procedure also had to account for three logistical complications: not all pupils would continue into the primary school attached to the nursery; pupils receiving the 30 week intervention had to attend nursery at the same time of day; less than 12 pupils were deemed eligible in 7 of the 34 nurseries selected. In order to address the first issue, pupils known to be moving to a different school were not included in the experiment and therefore the randomisation. However, destination was not known for all pupils and thus there was always a chance that a small number of pupils would drop out by moving to a different school. To account for the second and third problem, we devised a two-step procedure:

1. **Determining numbers in each group.** If there were exactly 12 pupils deemed eligible, 4 pupils were randomly allocated to each group. If there were 11 pupils, then 3 pupils were randomly allocated to the control group and 4 to each treatment group. If there were 10, then 3 pupils were randomly allocated to the control group and it was randomly determined whether T1 (30-week) or T2 (20-week) would have 3 or 4 pupils. If there were 9 pupils, then 3 were randomly allocated to each group.
2. **Determining T1 (30-week) pupils.** If at least a third of eligible pupils attended nursery in both morning and afternoon sessions, we randomly decided whether T1 pupils would be taken from the morning or afternoon session.⁹ However, if fewer than a third of the pupils attended a particular session, the selection of which session to allocate the T1 pupils to was deterministic. For example, if a total of 12 pupils were selected for the trial but only 2 pupils attended the morning session, T1 pupils would be drawn exclusively from the afternoon session. Eligible pupils attending the morning session would therefore have no chance of being allocated to T1.¹⁰

Random assignment should lead to small and statistically insignificant differences between each group in terms of mean age, gender shares and language composite scores. However, in any particular random draw it is possible that larger, significant differences can arise purely by chance. For example, one group might have a disproportionately large share of females. This may be particularly problematic in relation to subgroup analysis.

⁹ If the nursery had nine participating pupils, the afternoon and morning session required at least three pupils for random assignment.

¹⁰ If there were more than nine pupils in the nursery, we randomly assigned T1 pupils to the morning or afternoon session if there were at least four in each session. If there were exactly nine, then random assignment of pupils to the morning or afternoon session depended on at least three pupils in each session.

Pupils were therefore allocated to groups using a minimisation procedure, where an iterative process was used to determine the ‘optimal’ random assignment, which minimises the differences between treatment and control groups¹¹. The process outlined above was carried out, and then two diagnostic checks were performed. First, the three groups were compared to each other in terms of age, gender and language composite, and the number of statistically significant differences was recorded. Second, the difference in average characteristics between the three groups was calculated.

For each iteration, these two numbers were stored. The randomisation process was repeated 1,000 times, resulting in 1,000 different allocations. To identify the optimal randomisation, we first restricted our attention to the random assignments that led to zero significant differences between groups in terms of age, gender and language composite score. Among this set of assignments, we then selected the one that yielded the smallest value of the total differences in average characteristics. This was the final treatment allocation that we shared with the project team.

Table 3 presents the average characteristics for the three groups under the chosen random allocation. The groups are naturally similar in terms of the targeted variables (although this does not necessarily imply that the groups will be similar in terms of other characteristics, for example ethnicity or deprivation).

Table 3: Characteristics of each group

	30-week intervention	20-week intervention	Control Group
% female	49.2%	48.1%	48.8%
Average age in months	47.4	47.4	47.4
Average language composite score	-0.065	-0.073	-0.090

Analysis

Raw comparisons of post-test pupil test scores between treatment and control groups should provide unbiased estimates of the effect of the intervention if the randomisation has been successful. Methods that account for pupil characteristics will also yield unbiased estimates and in addition should be more precise estimates as a greater amount of the variation in test scores can be accounted for. In our analysis, we therefore present both raw comparisons and analysis that accounts for pupil characteristics and baseline test scores, with the latter representing our preferred estimates (both sets of estimates are shown in Table 11).

In particular, we control for gender, age, whether pupils have English as an additional language (EAL), whether pupils have known speech or language difficulties, and pre-treatment scores of all components of the language skills composite score. For the subsample of pupils with observed NPD records, additional characteristics that were controlled for are: whether pupils have been or currently are recorded as having special educational needs (SEN), whether pupils were eligible for free school meals (FSM) when in reception, ethnic group (minor categories included in the National Pupil Database), and deprivation of the pupil’s residential neighbourhood as measured by the IDACI percentile rank.

Our **preferred method** to account for the pre-test and pupil characteristics is Fully-Interacted Linear Matching (FILM) (Blundell *et al.*, 2005). FILM differs from standard Ordinary Least Squares (OLS) regression in that FILM linearly interacts the treatment effect with all pre-treatment characteristics and outcomes. This then provides an impact estimate for all individuals in the sample given their characteristics (for example, the estimated impact of treatment given they are male, have a low baseline test score, and have English as an Additional Language). We then calculate our preferred

¹¹ See Altman and Bland (2005) and Saghaei (2011) for discussions on the rationale behind minimisation.

impact estimate by averaging the impact estimates across all individuals in each treatment group, which corresponds to the average impact on those offered the treatment. As all outcomes are standardised by the pooled and unadjusted standard deviation prior to estimation, our impact estimates are always in effect size terms.¹²

The main advantage of FILM over OLS is that it is more flexible in allowing for the treatment effect to vary with individual characteristics. In our case, the FILM estimates are also generally more precise than both OLS and propensity score matching, which represents another advantage of FILM.

To account for the experimental design, standard errors are clustered at the school level to allow for correlation of pupil outcomes within schools. This approach is used across all methods presented in the paper. Another way to account for the experimental design in our analysis is to also allow pupil outcomes to explicitly depend on the school that they attend. This could take the form of a school effect that is assumed to be uncorrelated with all observable pupil characteristics (a random effects model) or one can explicitly estimate the individual effects of schools (a fixed effects model) (Wooldridge, 2010). Neither of these approaches should affect our estimates of the impact of the programme if the number of pupils in the treatment and control groups is equal across schools. However, estimating the treatment effects using these alternative methodologies represents another robustness check on our impact estimates (see Table C1). The random effects model is also identical to a hierarchical linear model with random intercepts (Raudenbush and Bryk, 2002).

The fact that the random assignment was re-run until balance was achieved has implications for the analysis which are still being debated. Both Bruhn and McKenzie (2009) and Scott *et al.* (2002) suggests that the most practical approach is to control for all covariates used in the randomisation, which we always do. Morgan and Rubin (2012) go further and show that standard errors calculated in the normal way are likely to be too conservative. They show that one can instead perform randomisation or permutation tests to perform inference, which are likely to generate smaller confidence intervals. However, these methods are still relatively new and only valid where a specific criterion has been used to determine acceptable randomisations (we instead chose the randomisation with the 'best' level of balance). We therefore still use conventional standard errors, but accept these are likely to be too conservative.

All analysis was undertaken on an intention-to-treat basis. In this context, this means we included pupils in the three schools that dropped out of the intervention. Not including such schools could threaten the balance of the treatment and control groups, particularly if schools dropped out in a non-random manner (for example, after perceiving low impact). However, it does mean that our common sample includes three schools where the interventions were not completed. This means that our impact estimates should be interpreted as including the effects of partial completion by three schools. It should be noted that we estimated the impacts again using only data from schools that completed the programme. The results (not reported here) were almost unchanged, with a change in the effect size of around 0.01.

To ensure that our effect estimates were not unduly sensitive to the choice of analysis method, **robustness checks** were performed by comparing treatment effect estimates across four alternative methodologies (raw comparison, OLS, FILM, and propensity score matching). As shown in Appendix C, treatment effect estimates were similar across all the different methodologies used, which supports the robustness of both the randomisation and the principal results reported here.

Sub-group analysis was also undertaken to examine whether the impact of the treatment varied between different groups of pupils. In particular, the impact on attainment among pupils without known speech and language difficulties or English as an Additional Language (EAL) was examined,

¹² In particular, controlling for additional characteristics only affects our estimated effect size if it affects the impact estimate. We do not use the residual variance after controlling for background characteristics to calculate the effect size.

as the process evaluation suggested these pupils were more receptive to the intervention. Treatment effects were also estimated for male and female pupils separately to identify any gender differences in the impact of the programmes. The results of this sub-group analysis are presented in Appendix B. We were able to undertake this sub-group analysis for the maximum number of pupils as data on speech and language difficulties, EAL status and gender was collected from schools directly.

To account for a wide range of pupil characteristics, NPD records were obtained for pupils whose schools had explicitly consented to this (those that had declined or did not respond were assumed not to consent). Because not all schools gave such consent, the analysis that considers these additional pupil characteristics is conducted on a subsample (239 out of the final common sample of 350). As FSM eligibility is recorded in the NPD, it was also possible to examine whether the intervention was more or less effective at improving the language skills of these relatively deprived pupils. This sub-group analysis was undertaken to conform to EEF requirements, but considerable differences in characteristics between the treatment and control groups compromise the validity of the results. All sub-group analysis was undertaken using FILM as the estimates are more precise and allow for a direct comparison with the estimated impact on the whole population of pupils.

Follow-up analysis was conducted using the same tests and composite outcomes as used for the post-test analysis. These tests were conducted about six months after the end of the 20- and 30-week interventions. By this time, schools had been offered the opportunity to deliver an alternative intervention (RALI) to the control group or receive a cash payment. According to the project team, 15 schools opted for the RALI intervention, 15 for the cash payments, and one did not respond. By the time of the follow-up test, eight schools had begun delivering the RALI intervention. No information was collected from schools as to how the cash payments were used. It is also not clear how many schools continued to deliver extra targeted support for treatment or control pupils (such as by continuing with elements of the Nuffield Early Language Intervention).

As a result, the interpretation of these longer-term treatment effects is relatively complex. Treatment and control pupils differed in terms of the interventions they received (Nuffield ELI or RALI, or cash to schools), the age at which they received them (start of reception or start of Year 1), and the time since the intervention took place (six months or straight after). We have decided to present the follow-up analysis, but the results are subject to the qualifications made in this section (the results are shown in Table 14).

All analysis was conducted using Stata 13 and undertaken on an intention-to-treat basis. This conforms to EEF guidance for interventions of this type. The syntax used is clearly documented and available to access from the UK data archive.

Implementation and process evaluation

The process evaluation was designed to explore views on the delivery, perceived impact and, to a lesser extent, implementation of the intervention. A qualitative case study approach was used to meet these objectives and involved in-depth interviews conducted in eight schools.

Sampling and recruitment for process evaluation

Schools were selected using a purposive sampling approach midway through the interventions.¹³ The goal was to achieve diversity across three key sampling criteria: geographical area, programme delivery mode, and whether schools experienced any delivery issues (including pupil attrition from programme and known fidelity issues¹⁴ such as deviation from the intended number or duration of

¹³ Selections were made in early 2014, and all interviews took place in February and March 2014. The interventions for the treatment groups were completed in April 2014.

¹⁴ Fidelity issues were identified by project team supporting schools to deliver the intervention. Fidelity issues include anything that deviates from the way the interventions were designed to be delivered according to the manuals in terms of both format and content.

sessions), as recorded by the project team. The selection of schools was also guided by secondary sampling criteria relating to pupil characteristics including number of pupils eligible for FSM and reported cases of SEN. Table 4 provides a breakdown of the achieved sample in relation to these primary sampling criteria.

While this sampling method ensured that the process evaluation recorded a range of experiences, it also means that the overall findings may not reflect the experiences of all schools involved. This is particularly the case given that the sample was constructed so that half of the schools recorded issues in the delivery of the programme, even though fewer than half of all schools participating in the trial may have experienced such problems. However, it was seen as beneficial for the programme development to include schools that had faced challenges so that these could be addressed in future.

Recruitment entailed a two-stage consent procedure. Advance letters and participant information sheets were sent to named programme coordinators in selected schools. The recruitment material provided information about the purpose of the study, what participation entailed, the timing of the fieldwork, and the voluntary and confidential nature of the study. Follow-up calls were then made to programme coordinators to ensure an understanding of the study, field any queries, and to gain consent for participation.

Table 4: Achieved sample - primary sampling criteria

Criteria	Details	Achieved sample
Geographical area	North	4 schools
	South	4 schools
Delivery mode	One teaching assistant delivering programme across nursery and reception	4 schools
	Two teaching assistants delivering programme (one in nursery, one in reception)	4 schools
Delivery issues	Delivery issues (fidelity ¹⁵ and/or pupil attrition)	4 schools
	No recorded delivery issues	4 schools

Once programme coordinators had agreed in principle for their schools to take part in the process evaluation, a second phase of consent commenced involving staff delivering the programme. This involved programme coordinators sending recruitment letters to relevant teaching staff to find out whether they were interested in participating in the study. If staff were interested, programme coordinators passed on contact details to NatGen researchers who then contacted staff to ensure they were happy to take part and to address any queries they had. Interviews were then arranged during these follow-up calls.

In total, twelve staff were interviewed across the eight case-study schools—nine teaching assistants and three senior teaching staff. This composition reflects the focus of the process evaluation on the views of those who were delivering the programme and had regular contact with children taking part in the intervention.

Conduct of the interviews

The interviews were designed to address the process evaluation research objectives described in the introduction in a way that minimised the burden of participation on schools. In view of this, the majority of the interviews were conducted over the telephone, with only four being conducted face-to-face. Interviews with staff delivering the programme (usually TAs) lasted up to one hour. Interviews with

¹⁵ These issues were recorded by the project team helping to support delivery of the interventions. They in turn indicated schools that had delivery issues in the sample frame.

senior staff (usually programme coordinators for the intervention) tended to be face-to-face and lasted no longer than 30 minutes.

All interviews were led by topic guides (included in Appendices E and F) that ensured systematic coverage of key issues while also allowing issues of relevance for individual respondents to be covered through detailed follow-up questioning. The interviews were digitally recorded and subsequently analysed using Framework, a systematic approach to qualitative data management developed by NatCen Social Research and now widely used in social policy research. Participants were assured their answers would be anonymised.

Timeline

The overall timeline for the interventions and evaluations is shown in Table 5 below. The project started with recruitment of schools by I CAN from September 2012. Screening of pupils was performed in January 2013. Training for the nursery component of the 30-week intervention was also done in early 2013. Pupils selected to participate were then randomised into one of the two treatment groups or the control group in April 2013. Soon afterwards, pre-test data was then collected from all pupils selected to participate and the 30-week intervention then commenced. In the autumn of 2013, TAs were trained in the reception component of the intervention and the 20-week intervention commenced soon afterwards. Post-test data was collected during the 2014 summer term from all participating pupils, after the interventions were complete. Finally, schools were offered the opportunity to deliver the alternative intervention (RALI) to pupils in the control group in the 2014 autumn term and follow-up test data was collected from all pupils by the end of December 2014.

Costs

To calculate the cost of the intervention we rely on information recorded by the project team regarding the monetary cost of the training and materials. We also take into account the indirect cost of the expected time commitments from staff. Monetary costs are presented in terms of the cost of training per staff member, as well their expected time commitment over the course of the trial. In addition, we indicate how many pupils this is expected to cover. Schools could deliver the intervention for smaller or larger numbers of pupils at their discretion, though there is no guarantee that the expected impact will be the same as that estimated in this evaluation.

This trial was commissioned before new guidance from the EEF on the systematic collection of cost data. All new trials are expected to collect such data in order to produce cost estimates in line with EEF guidance.

Table 5: Project Timeline

Date	Activity
September 2012	Programme development and school recruitment
January–April 2013	Teaching assistant training for nursery component
January–February 2013	Screening tests
March 2013	Preparation of online questionnaire to be sent to all schools to gather information about implementation of programme and other literacy strategies in operation
April 2013	Randomisation performed based on screening data received from the project team
March–April 2013	Pre-test data collected
April 2013	30-week intervention begins in nurseries
May–July 2013	Delivery team invited schools to complete online questionnaire to gather information about implementation of programme and other literacy strategies in operation
Autumn 2013	Teaching assistant training for reception component
September 2013	20-week intervention begins
September 2013–March 2014	Analysis of responses to online questionnaires; contact eight schools to invite them to participate in further qualitative research; carry out face-to-face/telephone interviews with TAs, teachers/coordinators
April–June 2014	End of both the 30- and 20-week interventions
May–July 2014	Post-test data collected
September 2014	Schools offered chance to deliver RALI to control group pupils or receive a cash payment
November 2014	Post-test data received from project team
Oct–Dec 2014	Follow-up test data collected
March 2015	Follow-up test data received from project team
End June 2015	Final evaluation report delivered to the EEF

Impact evaluation

Participants

As previously indicated, the intervention was targeted at primary schools with attached nurseries. Within each of the 34 participating nurseries, pupils were then screened for eligibility for the intervention. The parents of the children with the 12 lowest composite language scores were invited to give consent for their child to participate in the trial. If the parents of any of these children opted out of the trial, the intention was to approach parents of the next lowest scoring child up to a maximum eligibility score. However, as indicated in the randomisation section, less than 12 pupils in some schools were sometimes deemed eligible and given parental permission. In total, consent was successfully obtained for 394 children.

Table 6 compares the number of pupils in each group at the time of randomisation with the final sample used in the evaluation. Figure 3 provides a flow chart (a slightly adapted version of the CONSORT flow chart) showing how the number of schools and pupils in the intervention were determined at each stage.

Because of the method used to identify eligible schools, most of the 394 pupils progressed into participating primary schools and completed post-intervention testing. However, destination primary school was not known for all pupils at the time of randomisation and some pupils dropped out of the trial as a result of moving to a non-participating school. In addition to this, ten pupils had to be excluded from the impact evaluation due to missing data, three of whom did not complete all of the pre-tests and seven of whom did not complete (almost all) of the post-tests.

Table 6: Number of pupils at randomisation and in final sample

Group	At randomisation	Complete treatment	Complete treatment (no missing data)
All pupils in trial	394	360	350
30-week treatment	132	117	114
20-week treatment	133	124	121
Control	129	119	115

Figure 3: Flow of participants at different stages of the trial

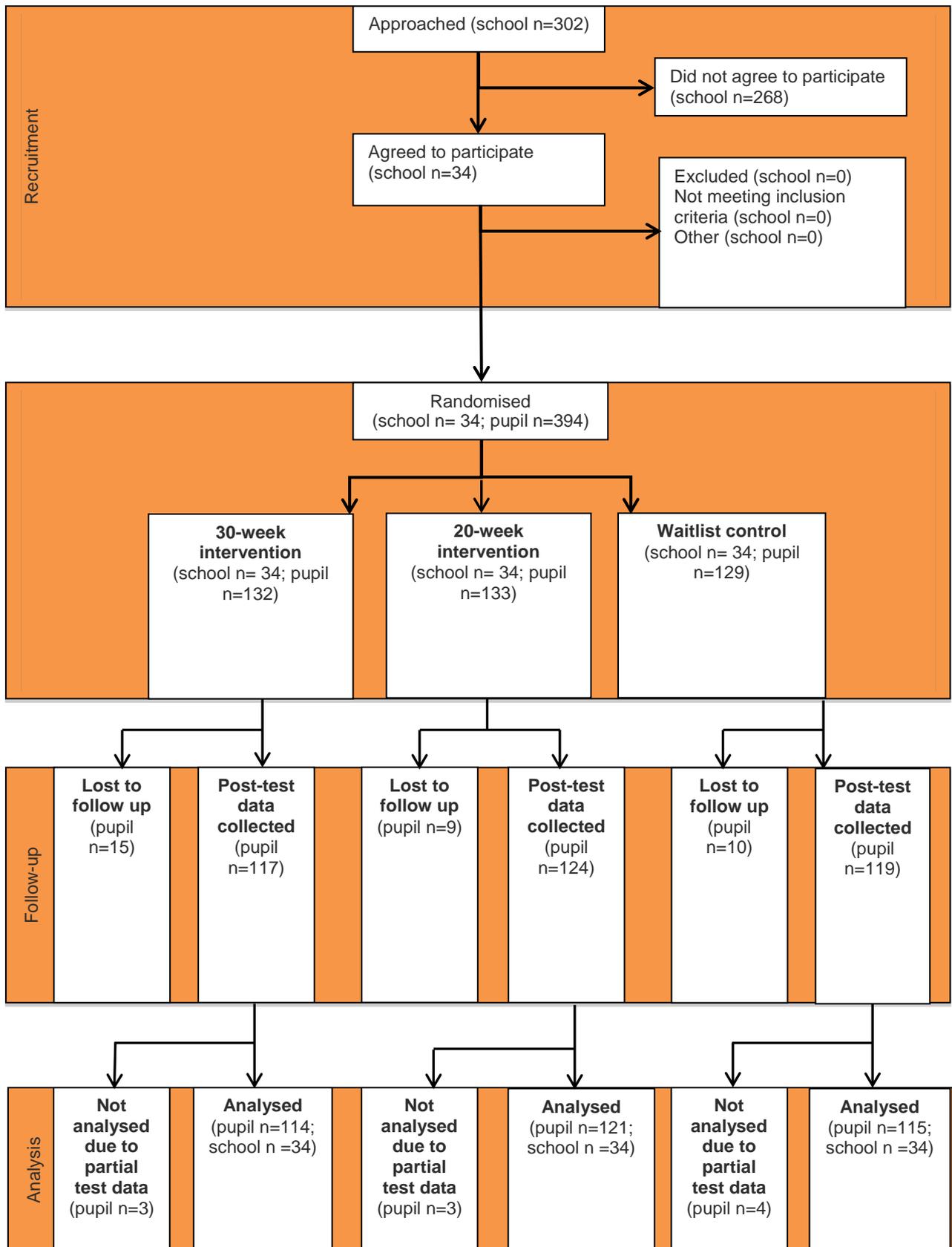


Table 7 shows how the minimum detectable effect size changed across different stages of the trial. At the planning stage, the project team aimed to recruit 360 pupils across 30 schools. At this point we assumed that pre-test characteristics and outcomes would explain only 20% of the variation in the post-test outcome. Using other standard assumptions, this would give a minimum detectable effect size of 0.32. In reality, the project team managed to recruit 4 more schools than expected, giving a total sample size of 394 and a minimum detectable effect size of 0.31.

Table 7: Minimum detectable effect size at different stages

Stage	Number of pupils (T1, T2, C)	Correlation between baseline characteristics & post-test	ICC	Randomisation method	Power	Alpha	Minimum detectable effect size (MDES)
Planning	360 (120, 120, 120)	(predicted) 0.2	n/a	within school	80%	0.05	0.32
Randomisation	394 (132,133,129)	(predicted) 0.2	n/a	within school	80%	0.05	0.31
Analysis	350 (114, 121, 115)	(actual) 0.55	n/a	within school	80%	0.05	0.25
NPD Analysis	239 (80, 83, 76)	(actual) 0.59	n/a	within school	80%	0.05	0.29

We were not able to include all pupils in the final analysis, mostly due to pupils moving school between nursery and primary school. However, pre-test characteristics and outcomes explained much more of the variation in post-test outcomes than expected (about 55%). As a result, the minimum detectable effect size at the analysis stage was much lower at 0.25.

Not all schools provided Unique Pupil Numbers to enable linkage to the National Pupil Database. As a result the sample size for the analysis using the NPD is smaller (239 pupils), though we are able to explain a slightly greater share of the variation in post-test outcomes as a result of having more background characteristics. This gives a minimum detectable effect size of about 0.29.

Pupil characteristics

As the trial involved a select sample of schools, it is important to recognise that the impact of either intervention in other schools may differ from the results presented here. Table 8 presents key characteristics that may be expected to influence the impact of the programmes for those schools in the trial and for all primary schools in England.¹⁶ It shows that, compared to the average English primary school, schools involved in the trial had, on average, significantly more pupils, and a significantly lower ratio of pupils to full-time equivalent TAs. The latter result is important as it may be more difficult to deliver the programme where the ratio of pupils to TAs is higher, particularly in light of the delivery problems highlighted in the process evaluation. Schools in the trial also had a significantly greater proportion of pupils eligible for FSM, which is to be expected given that disadvantaged schools were explicitly targeted. If school-wide deprivation impacts the effectiveness of the intervention in some way, then treatment effects may also vary in any future implementation of the programme in less deprived schools. However, there were no statistically significant differences between the trial schools and primary schools in general in respect of four key factors: pupil to

¹⁶ Some staffing variables were missing due to changes in school identifiers and missing data. Key Stage 2 results are also clearly missing for schools with no pupils in year 6 (e.g. infant schools).

teacher ratio, current levels of performance at Key Stage 2, proportion of SEN pupils, and the proportion of academies.

Table 8: Comparison of trial schools with all primary schools in England

Variable	All primary schools		Trial Schools		Differences	
	n	Mean (sd)	n	Mean (sd)	Actual	Effect Size
School-level continuous variables						
Number of pupils	16,613	255.93 (143.74)	34	386.35 (149.04)	130.42***	0.907
Pupil:Teacher Ratio (FTE)	16,515	20.65 (5.47)	34	21.56 (4.21)	0.91	0.167
Pupil:Assistant Ratio (FTE)	16,467	32.01 (19.42)	34	27.27 (10.48)	-4.74**	-0.244
School-level categorical variables						
Proportion of pupils achieving expected level at Key Stage 2 in English and Maths	14,152	0.80 (0.13)	29	0.77 (0.08)	-0.03	-0.2
Proportion of pupils eligible for FSM	16,613	0.17 (0.14)	34	0.26 (0.10)	0.09***	0.696
Proportion of pupils with SEN	16,613	0.01 (0.02)	34	0.02 (0.03)	0.01	0.52
Proportion of pupils with EAL	16,613	0.14 (0.21)	34	0.24 (0.19)	0.11***	0.493

Note: * indicates that the difference in means is significant at the 10% level ** at the 5% level *** at the 1% level. Standard deviations are reported in brackets. Data for all school characteristics relates to January 2014 and was downloaded from the Department for Education 2014 Performance Tables (http://www.education.gov.uk/schools/performance/download_data.html).

The group average characteristics of the 350 pupils with complete pre- and post-intervention test records (the common sample) are presented in Table 9. This includes characteristics considered at randomisation, other pupil characteristics (including variables collected from the NPD where permission was given), the primary pre-test outcome and its components, and other baseline pre-test scores (as detailed in the outcomes measures subsection above). We show the mean and standard deviation of each characteristic within each group and the difference relative to the control group expressed as an effect size (calculated by dividing the actual difference by the pooled standard deviation across the common sample). The overall sample size for each characteristic is then shown on the right hand side. For most characteristics, this represents the 350 pupils in the common sample. For data collected from the NPD, this is reduced to the 239 pupils where a link was possible.

In general, the results show that, despite the reduction in sample size due to pupil dropout and missing post-intervention data, the groups remain balanced in terms of the characteristics that were considered during the randomisation. Pupils who received the 30-week intervention had, on average, slightly higher scores for the screening language composite but the difference is not statistically significant and equivalent to an effect size of around 0.1.

The groups are also well-balanced in terms of other pupil characteristics that were not included in the randomisation process. There are no statistically significant differences across groups and the differences that do exist are generally small in absolute value. The only exceptions are that the 20-week group does appear to be more deprived than the control group in terms of the proportion of pupils eligible for FSM (34% compared with 24%, equating to an effect size of 0.22), and fewer pupils seem to have special educational needs and fewer are of White-British ethnicity. Although these additional variables are only observed for 238 pupils, it is possible that these imbalances could exist in the sample as a whole. Fortunately, our estimates of the impact of the two treatments are very similar for the whole sample, and for the sub-sample of those we can link to the NPD, when we control for additional covariates available in the NPD (see Table 13). This suggests that any small

differences between the treatment and control groups are not too problematic. These inter-group differences highlight the importance of accounting for pupil characteristics when estimating treatment effects.

The differences between treatment and control groups are also small in terms of the primary pre-test language composite outcome, as well as its components. The differences between the groups all equate to an effect size of 0.1 or smaller. Differences are equally small in terms of other pre-test scores collected by the project team, with the one exception of a slightly larger difference on the British Picture Vocabulary Scale.

The overriding conclusion from this analysis is that the groups are largely balanced in terms of pre-test and baseline characteristics. There are a few small differences, particularly in terms of characteristics collected from the NPD; however some differences are to be expected given the relatively small sample sizes. Furthermore, none are statistically significant and our impact estimates are largely unchanged when we control for additional characteristics from the NPD.

Table 9: Comparison of baseline characteristics

Variable	30-week (T1)		20-week (T2)		Control	Total N
Characteristics considered at randomisation	Mean (sd)	T1 - C (as effect size)	Mean (sd)	T2 - C (as effect size)	Mean (sd)	
Proportion of pupils who are female	0.49 (0.50)	0.01	0.48 (0.50)	-0.02	0.49 (0.50)	350
Age in months	46.10 (0.50)	-0.08	46.10 (0.50)	-0.04	46.23 (0.50)	350
Screening language composite	0.00 (0.50)	0.11	0.01 (0.50)	0.08	-0.06 (0.50)	350
Other pupil characteristics						
Proportion of pupils with EAL	0.15 (0.36)	-0.02	0.18 (0.39)	0.12	0.14 (0.35)	350
Proportion of pupils with known speech and language difficulties	0.03 (0.17)	0	0.04 (0.20)	0.09	0.03 (0.16)	350
Proportion of pupils eligible for FSM	0.28 (0.45)	0.03	0.34 (0.48)	0.22	0.24 (0.43)	239
Proportion of pupils with SEN (statement or school action plus)	0.13 (0.34)	-0.02	0.10 (0.30)	-0.18	0.16 (0.37)	239
Proportion of pupils of White-British Ethnicity	0.66 (0.47)	-0.07	0.63 (0.49)	-0.15	0.70 (0.46)	239
Primary pre-test outcomes (standardised)						
Language composite	0.00 (1.00)	0.03	0.00 (1.09)	0.02	-0.02 (0.94)	350
CELF expressive vocabulary	0.00 (1.00)	0.06	0.01 (1.07)	0.04	-0.03 (0.98)	350
APT information	-0.00 (1.00)	0.04	-0.04 (1.00)	-0.04	-0.00 (0.99)	350
APT grammar	-0.00 (1.00)	0.1	-0.06 (1.02)	-0.04	-0.02 (0.94)	350
Listening comprehension	0.00 (1.00)	-0.09	0.09 (1.15)	0.1	-0.00 (0.86)	350
Other pre-test outcomes (standardised)						

CELF sentence structure	0.00 (1.00)	0.08	-0.00 (1.01)	0.04	-0.04 (1.01)	350
Vocabulary naming	-0.00 (1.00)	0.01	0.03 (0.99)	0.06	-0.02 (0.99)	350
Vocabulary definitions	-0.00 (1.00)	-0.1	0.04 (1.06)	0	0.03 (0.91)	350
British picture vocabulary scale	0.00 (1.00)	0.02	0.08 (1.03)	0.14	-0.05 (1.06)	350
Total Sample Size	114		121		115	350
Total Sample with NPD data	80		83		76	239

Note: * indicates that the difference in means (TX - C) is significant at the 10% level ** at the 5% level *** at the 1% level. Standard deviations are reported in brackets.

In addition to the main results presented in Table 11, which have been estimated using the final sample of pupils with complete post-intervention test records, treatment effects are also reported for several subsamples to examine whether the impact of the interventions varies (Table 13). We have thus also compared pupil characteristics across treatment and control groups in these subsamples (results shown in Appendix B).

For the subsample of pupils who do not have known speech or language difficulties or EAL and the subsample of pupils with observed NPD records, the differences observed between treatment and control pupils are broadly similar to those in the full sample. For the male and female subsamples, the trial groups remain relatively well-balanced in terms of most baseline characteristics, but the presence of some statistically significant differences mean the results for these subsamples are less secure than the estimates computed using the full sample.

The subsample of FSM eligible pupils displays several notable differences between the treatment and control groups in terms of the pre-test language skills measure, ethnicity, and special educational needs. Although many of these differences are not statistically significant, they are large in absolute value and compromise the extent to which results from the randomised controlled trial can be used to accurately identify the impact of the treatments on FSM eligible pupils. As this subsample is of particular interest to the EEF, future trials should consider including FSM eligibility in the randomisation process to allow impacts to be analysed with greater confidence.

Outcomes and analysis

We start by showing the raw differences in our primary post-test outcomes before presenting our main impact estimates of the Nuffield Early Language Intervention. Table 10 shows the average level of the primary and secondary post-test outcomes, and their components, for both treatment groups and the control group. All outcomes are standardised within the common sample so that the difference between treatment and control groups can be interpreted as an effect size.

This shows that, based on the raw outcomes, both treatment groups showed evidence of better language skills than the control group (with this difference statistically significant at the 5% level for the 30-week intervention, and at the 10% level for the 20-week version). There were smaller differences between treatment and control groups in terms of word-level literacy skills (and, in all but one case, not statistically significant).

Table 10: Comparison of post-test outcomes

Variable	30-week (T1)		20-week (T2)		Control	Total N
Primary post-test outcomes (standardised)	Mean (sd)	T1 - C (as effect size)	Mean (sd)	T2 - C (as effect size)	Mean (sd)	
Language composite	0.11 (0.97)	0.28**	0.05 (1.03)	0.22*	-0.17 (0.98)	350
CELF expressive vocabulary	0.09 (0.97)	0.25*	0.06 (1.01)	0.22*	-0.16 (1.01)	350
APT information	0.03 (1.00)	0.12	0.06 (0.98)	0.15	-0.09 (1.02)	350
APT grammar	0.14 (0.94)	0.29**	0.01 (1.04)	0.16	-0.15 (1.00)	350
Listening comprehension	0.08 (0.97)	0.2	0.03 (1.09)	0.15	-0.12 (0.93)	350
Secondary post-test outcomes (standardised)						
Word-level literacy composite	-0.01 (0.74)	0.07	0.09 (0.97)	0.17	-0.08 (0.93)	350
YARC letter sound knowledge	0.06 (0.76)	0.12	0.01 (1.11)	0.07	-0.06 (1.09)	350
YARC early word reading	-0.05 (0.90)	0.01	0.10 (1.07)	0.16	-0.06 (1.02)	350
Spelling total	-0.04 (0.90)	0.09	0.15 (1.03)	0.28**	-0.13 (1.05)	350
Total Sample Size	114		121		115	350

Note: * indicates that the difference in means (TX - C) is significant at the 10% level ** at the 5% level *** at the 1% level. Standard deviations are reported in parentheses. All outcomes are standardised within the common sample to have a mean of zero and standard deviation of one.

Table 11 presents our main estimates of the impact of the Nuffield Early Language Intervention on the primary outcome (a composite of language skills) and on the secondary outcome (a composite of word-level literacy skills). The table starts by showing the average level of the raw outcomes across the 20-week and 30-week intervention groups as well in the control group (all standardised to have a mean of zero and a standard deviation of one within the common sample). We also show the sample sizes these are based on and the number of missing observations as compared with the total number of non-missing observations on each outcome. The missing observations result from imposing a common estimation sample (in other words, dropping cases where one or more of the pre-test or post-test outcomes are missing).

The differences in raw outcomes will provide unbiased treatment effect estimates if the covariates are well balanced across groups. However, our preferred impact estimates control for the baseline differences in pupil characteristics and pre-test scores. Where the covariates are well balanced, controlling for them in our specifications reduces the variance of our estimates and therefore increases statistical power. Controlling for these covariates will also address any potential imbalances caused by pupil drop out, or characteristics not considered at randomisation. On the right hand side of each table, we show the estimated impact based on our preferred methodology accounting for baseline characteristics (FILM). Also shown are the sample sizes involved and the p-value for a two-sided test for a null hypothesis that the estimated impact is zero.

We estimate that both versions of the Nuffield Early Language Intervention have large, positive and statistically significant impacts on the composite measure of language skills. Our preferred estimates suggest that the 30-week intervention improved pupils' language skills by the equivalent of 0.27

standard deviations. This is very similar to the difference in raw outcomes between the treatment and control group (0.28), as we might expect given the very small differences in baseline characteristics. The 20-week intervention is estimated to improve language skills by 0.16 standard deviations, though is only statistically significant at the 10% level. This is below that implied by the difference in raw outcomes (0.22 standard deviations), though not substantially so.

Although the impact estimate for the 30-week intervention, starting in nursery school, was larger than the estimate for the 20-week intervention, delivered in schools only, it is not possible to say with confidence that the 30 week intervention is more effective. This is because the difference between the 30-week intervention and the 20-week intervention is not statistically significant: this may be because there is no difference between the respective impacts, or because the trial was not sufficiently powered to detect a difference.

In contrast to the impact on language skills, the second panel of results in Table 11 suggest that the impact of each intervention on word-level literacy skills is small and not statistically significant. These findings are consistent with an earlier randomised controlled trial of the 30-week Nuffield Early Language Intervention which identified significant improvements in oral language skills and a weaker impact on word-level literacy skills.¹⁷

Therefore, the results suggest a positive impact of the Nuffield Early Language Intervention on language skills, but little impact on word-level literacy skills. At this point, however, it is important to remember that the randomisation of pupils within schools does create the potential for spillover effects on the control group. These could be positive if TAs used the intervention materials with control pupils. We think this is unlikely as the training emphasised the importance of the experimental conditions and the process evaluation confirmed that TAs were aware of this. Negative spillovers could occur if the intervention led to high workloads amongst teaching assistants and control pupils lost out as a result. The process evaluation does report higher than expected workloads among TAs. This suggests potential for negative spillovers. However, the size of any spillover effect is uncertain.

Table 11: Impact of Nuffield Early Learning Intervention on language and literacy skills

Outcome	Raw means				n in model (treatment, control)	Effect size Estimated Impact (95% CI)	p-value
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)			
Language skills (primary outcome)							
30-week treatment	114 (3)	0.11 (-0.07; 0.29)	115 (4)	-0.17 (-0.35; 0.01)	229 (114, 115)	0.267*** (0.073; 0.461)	0.007
20-week treatment	121 (3)	0.05 (-0.13; 0.24)	115 (4)	-0.17 (-0.35; 0.01)	236 (121, 115)	0.161* (-0.016; 0.339)	0.075
Word-level literacy skills (secondary outcome)							
30-week treatment	114 (3)	-0.01 (-0.17; 0.14)	115 (4)	-0.09 (-0.29; 0.10)	229 (114, 115)	0.062 (-0.159; 0.283)	0.582
20-week treatment	121 (3)	0.10 (-0.10; 0.30)	115 (4)	-0.09 (-0.29; 0.10)	236 (121, 115)	0.127 (-0.082; 0.337)	0.234

Note: * indicates that the treatment effect is significant at the 10% level ** at the 5% level *** at the 1% level. Impact estimates are calculated using Fully Interacted Linear Matching (FIM). Covariates included are: age; gender; EAL status; known speech and language difficulty status; standardised language composite baseline score; standardised baseline scores for components of language composite; standardised baseline scores for CELF sentence structure, vocabulary naming, vocabulary definitions and the British picture vocabulary scale.

¹⁷ Fricke *et al.* (2013).

Robustness checks

We performed a number of robustness checks to confirm these findings. First, we find that the impact estimates are largely unchanged when the sample is restricted to pupils with observed NPD records and we account for additional characteristics such as FSM eligibility and ethnicity.¹⁸ These estimates are shown in Table 13 below, which also incorporates our sub-group analysis.

To examine the robustness of the FILM estimates reported here, treatment effects were also computed using a range of alternative methods. Table C1 shows that the estimated effects on the language composite have a similar magnitude across the different methodologies, which supports the robustness of the results presented here. In several cases, the FILM estimates have a higher level of statistical significance than the alternative approaches, which is a result of the greater precision of FILM estimation. It should also be noted that alternative ways of accounting for the experimental design (random effects and fixed effects) produce very similar estimates of the impact of the interventions and very similar standard errors. This further supports the robustness of our results.

The impact estimates presented here show that both treatments had a positive and statistically significant effect on the language skills of pupils, as measured by a language skills composite measure. To improve the interpretation of this finding, it is useful to examine how treatment effect estimates vary between the different component scores of the language skills composite. Table 12 presents this information for the entire sample of pupils. This shows that the 30-week treatment had a large and statistically significant impact on all components of the language skills composite except the Action Picture Test (APT) information score, with the largest impact on the APT grammar score. Although the 20-week treatment only has a significant effect on the language skills composite, rather than its components, the largest positive treatment effect is also on the APT grammar score.

Table 12: Language composite score and component treatment effect estimates—all pupils

Outcome	30-week treatment			20-week treatment		
	Treatment effect	95% Confidence Interval	n	Treatment effect	95% Confidence Interval	n
Language composite (primary outcome)	0.267***	(0.073; 0.461)	229	0.161*	(-0.016; 0.339)	236
CELF expressive vocabulary	0.245***	(0.061; 0.430)	229	0.13	(-0.096; 0.356)	236
APT information	0.09	(-0.138; 0.319)	229	0.138	(-0.049; 0.325)	236
APT grammar	0.282**	(0.061; 0.503)	229	0.151	(-0.068; 0.370)	236
Listening comprehension	0.202*	(-0.016; 0.421)	229	0.076	(-0.128; 0.280)	236

*Note: * indicates that the treatment effect is significant at the 10% level ** at the 5% level *** at the 1% level.*

Standard deviations are reported in square brackets.

Covariates included are: age; gender; EAL status; known speech and language difficulty status; standardised language composite baseline score; standardised baseline scores for components of language composite; standardised baseline scores for CELF sentence structure, vocabulary naming, vocabulary definitions and the British picture vocabulary scale.

Sub-group analysis

To examine whether the treatments were more or less effective for certain groups of pupils, treatment effects were also estimated using several subsamples that were defined by particular characteristics. These results are presented in Table 13. Although instructive, it should be noted that these results are

¹⁸ Using the subsample of pupils with observed NPD records but without including extra pupil characteristics returned treatment effect estimates that were very similar to those calculated using the entire sample of pupils.

less secure than the treatment effects in Table 11 owing to the inter-group differences shown in Appendix B and discussed above.

The results in Table 13 suggest both treatments are more effective at improving the language skills of pupils without known speech and language difficulties or EAL. This is consistent with the process evaluation which finds that the interviewed teaching assistants believed EAL pupils and pupils with known speech and language difficulties were less receptive to the interventions. There are also clear gender differences in the impact of treatments on the language composite score. The estimates suggest that the 30-week treatment is more effective at improving the language skills of female pupils while the 20-week intervention is more effective at improving the language skills of male pupils.

Interestingly, the estimated treatment effects suggest both treatments are considerably less effective for pupils eligible for FSM: the treatment effect estimates for this subgroup are very large and negative, albeit insignificant. These counter-intuitive results may be explained by the imbalances between intervention groups, as discussed above. The combination of the subsample imbalances and the lack of a statistically significant effect cast further doubt on the ability of the trial to accurately reveal the effectiveness of the treatments on FSM eligible pupils.

Table 13 also presents the estimated treatment effects of the two interventions on the word-level literacy skills composite. Again, these results suggest neither treatment has a significant impact on word-level literacy skills. Although the 20-week treatment appears to have a positive significant effect on male pupils, the treatment effect estimate is only significant at the 10% level, and is less secure than the results shown in Table 10 owing to inter-group differences for the male subsample, as illustrated in Appendix B Table B2.

Table 13: Subsample treatment effect estimates

30-week treatment				20-week treatment		
Language composite score (primary outcome)						
Sample	Treatment effect	95% Confidence Interval	n	Treatment effect	95% Confidence Interval	n
Pupils with observed NPD records [†]	0.295**	(0.041; 0.549)	156	0.169	(-0.134; 0.462)	159
Pupils without known speech and language difficulties or EAL	0.301***	(0.115; 0.486)	192	0.260**	(0.029; 0.491)	190
Male pupils	0.117	(-0.146; 0.379)	117	0.202*	(-0.020; 0.424)	122
Female pupils	0.374**	(0.054; 0.695)	112	0.102	(-0.143; 0.348)	114
Pupils eligible for FSM [†]	-0.944	(-2.101; 0.214)	38	0.348	(-0.443; 1.140)	46
Word-level literacy composite score (secondary outcome)						
Sample	Treatment effect	95% Confidence Interval	n	Treatment effect	95% Confidence Interval	n
Pupils with observed NPD records [†]	0.015	(-0.272; 0.301)	156	0.051	(-0.248; 0.351)	159
Pupils without known speech and language difficulties or	0.091	(-0.164; 0.345)	192	0.18	(-0.054; 0.414)	190

EAL						
Male pupils	0.236	(-0.143; 0.616)	117	0.324*	(-0.043; 0.690)	122
Female pupils	0.02	(-0.197; 0.237)	112	0.001	(-0.320; 0.323)	114
Pupils eligible for FSM [†]	-1.067	(-3.137; 1.002)	38	-0.259	(-1.234; 0.715)	46

*Note: * indicates that the treatment effect is significant at the 10% level ** at the 5% level *** at the 1% level. Standard deviations are reported in square brackets. Covariates included are: age; gender; EAL status; known speech and language difficulty status; standardised language composite baseline score; standardised baseline scores for components of language composite; standardised baseline scores for CELF sentence structure, vocabulary naming, vocabulary definitions and the British picture vocabulary scale. † Indicates additional covariates from the NPD data: eligibility for FSM when in reception; ethnicity; whether the pupil has ever received SEN support; and IDACI rank percentile.*

Follow-up tests

Pupils' language and word-level literacy skills were also tested with a follow-up six months after the end of the main intervention phase. By this stage, schools had been offered the opportunity to deliver an alternative intervention (the Reading and Learning Intervention, RALI) or receive a cash payment to deliver another intervention of their choice. According to the project team, about 15 schools opted for the RALI intervention, of which about eight had begun delivery by the time of the follow-up test. Fifteen schools opted for the cash payment and one school did not respond.

As a result, differences between groups will therefore measure the relative effects of the completed treatment interventions and the part-completed alternative intervention offered to control pupils (which will have started, but are unlikely to have been completed). Even this interpretation is further complicated by the other differences between the treatment and control conditions: the control group will have received any alternative treatment at a later stage; and, six months has elapsed since the end of the original interventions. With these qualifications in mind, Table 14 shows the estimated differences between the two original treatment groups and the original control group (both for the repeated primary language composite outcome and the secondary word-level literacy composite).

Interestingly, the differences between the treatment and control groups in terms of the primary language composite outcome have actually slightly increased in size by the time of the follow-up test as compared with the immediate post-test outcomes. The difference between the 30-week intervention and the original control group increased from 0.27 to 0.37 standard deviations, and the difference between the 20-week treatment group and the control group increased from 0.16 to 0.21 standard deviations. A similar increase is seen for pupils with NPD data. However, there remains no statistically significant difference in terms of word-level literacy scores.

What could be driving this increase in the estimated effects of the original treatment compared with the original control group? First, we must acknowledge that the control group have been offered and many will have begun to receive an alternative intervention, though delivered at a later age. However, this difference can only explain the increase in the estimated impact of the original interventions if the control group intervention had a negative impact. Although possible, we think this is unlikely as previous evidence suggests this intervention has previously had positive impacts. Another potential explanation is that the age at which intervention is received matters and that earlier intervention has a larger effect. However, for this to be the main explanation, it would need to be the case that intervention starting at age three to four has the potential to have a large effect, whilst intervention around age five has the potential to have a near zero effect. Such a dramatic change seems implausible, though not impossible.

In our view, a more plausible explanation is the time elapsed since the intervention. These results are consistent with the original interventions having given children the tools to further improve their language skills after the end of the intervention. For example, the process evaluation found that TAs thought pupils' confidence had improved as a result of the intervention. This, and their improved

language skills, may have enabled pupils to access other areas of the curriculum after the intervention. They are also consistent with TAs continuing to use the intervention with treatment pupils and thus increasing the dosage beyond the stated 20 or 30 weeks, which they were entitled to do if they so decided. Unfortunately, although a number of TAs reported to the process evaluation that they planned to continue using the intervention materials after it had finished it is not clear exactly how many actually did.

Having said this, it is important not to over-interpret this finding as the confidence intervals surrounding all the impact estimates are relatively wide. However, the results are suggestive of an increase in the treatment effect between the end of the intervention and the six-month follow up.

Table 14: Follow-up test impact estimates

30-week treatment				20-week treatment		
Follow-up language composite score (primary outcome)						
Sample	Treatment effect	95% Confidence Interval	n	Treatment effect	95% Confidence Interval	n
All pupils	0.367***	(0.170; 0.563)	229	0.211**	(0.015; 0.407)	236
Pupils with observed NPD records including NPD covariates	0.464***	(0.189; 0.739)	156	0.265*	(-0.008; 0.538)	159
Follow-up word-level literacy composite score (secondary outcome)						
Sample	Treatment effect	95% Confidence Interval	n	Treatment effect	95% Confidence Interval	n
All pupils	0.054	(-0.259; 0.366)	229	0.066	(-0.151; 0.283)	236
Pupils with observed NPD records including NPD covariates	-0.05	(-0.332; 0.233)	156	-0.012	(-0.425; 0.401)	159

*Note: * indicates that the treatment effect is significant at the 10% level ** at the 5% level *** at the 1% level. Covariates included are: age; gender; EAL status; known speech and language difficulty status; standardised language composite baseline score; standardised baseline scores for components of language composite; standardised baseline scores for CELF sentence structure, vocabulary naming, vocabulary definitions and the British picture vocabulary scale. Additional NPD covariates are: eligibility for FSM when in reception; ethnicity; whether the pupil has ever received SEN support; and IDACI rank percentile.*

Cost

The main costs of the intervention relate to training, materials, and the time of teaching assistants to deliver the programme.

In order to deliver the 30-week intervention, schools would need to pay just under £2,500 to cover training and materials for one person (assumed to be a TA) to deliver the nursery and reception components (this need not be the same person delivering both components though). This includes:

- a one day training event for the nursery component (£800 per person);
- nursery component materials and manuals (£250 per pack);
- a two-day training event for the reception component (£1,000 per person);
- reception component materials and manuals (£400 per pack); and
- optional materials (Nuffield Bear £10, book pack £19).

The cost of training a staff member to deliver the 20-week intervention would be around £1,400 (the two day training course plus materials).

Staff time is also a significant cost. For the nursery component, nursery staff were expected to deliver three 20-minute sessions per week for four pupils for ten weeks, plus one hour of preparation per week. A commitment, therefore, of two hours per week for ten weeks—a total of 20 hours for each staff member delivering the intervention.

For the reception component, teaching assistants were expected to spend 4.5 hours per week delivering the programme for 20 weeks to four pupils: three 30-minute group sessions totalling 1.5 hours; two 15-minute individual sessions per child per week (two hours); plus one hour of preparation time per week. This gives a total required time commitment of 90 hours.

Considered together, to enable one staff member to deliver the 20-week intervention to four pupils, schools would need to pay £1,400, and this staff member would need to spend 90 hours preparing and delivering the intervention spread over 20 weeks. In comparison, schools would need to pay £2,500 to have a member of staff capable of delivering the 30-week version and this would require a total of 110 hours spread over 30 weeks.

It should be noted that these calculations all assume four pupils receiving the intervention from one staff member (as it was in the trial). Schools could clearly choose to deliver the intervention to groups of more or fewer than four children, though we do not know whether this would increase or lessen the impact of the intervention.

These calculations exclude any further staff time or disruption costs. For instance, some staff time is likely to be required to coordinate and organise the delivery of the intervention, and the process evaluation emphasised some disruption costs encountered by taking pupils out of lessons. Furthermore, the staff time detailed above is likely to be an under-estimate of the true level of staff time required as further preparation and organisation time is likely to be required (for example, the process evaluation noted that some staff reported finding it hard to deal with the additional workload caused by delivering the treatments).

The table in the executive summary presents the EEF cost rating for each intervention. This is based on the *additional monetary cost* to a school of delivering the interventions. It does not include, for example, the costs of staff time for staff already employed in the school. The cost is calculated as the cost per pupil over a three year period, and it is therefore necessary to make an assumption about how many children would receive the intervention each year. Based on the staff time requirements, one full-time teaching assistant could, in theory, deliver the intervention to up to eight groups of four children each year. However, it is unlikely that a school would have this many children in nursery/reception class who were suitable for the intervention. For the trial, it was decided that the twelve children with the lowest composite language scores would be eligible. The cost rating is therefore based on the assumption that twelve children in a given school (three groups of four) will receive the intervention each year. The EEF cost ratings are explained in Annex G.

Summary

The estimates presented here suggest that both the 30-week and 20-week Nuffield Early Language Interventions, as delivered by I CAN, had a significant positive effect on a composite measure of language skills, with the effect larger for the 30 week intervention. There is evidence that both treatments had a greater positive impact on pupils without known speech and language difficulties or EAL, which is consistent with the findings of the process evaluation. Neither treatment appears to have had a substantive impact on a composite measure of word-level literacy, in common with trials of earlier versions of this intervention. Interestingly, the differences between the language skills of

pupils in the treatment and control groups increased at a follow-up test six months after the end of the original interventions, even though the control group had received an alternative intervention by this time. This provides some evidence to suggest that the effects of the intervention may increase over time.

Process evaluation

The process evaluation explored professionals' experiences of delivery and views on perceived impact. This included views on the implementation and setting up of the programme, the delivery approach in schools, perceptions about impact and features critical to this and views on the sustainability of the programme. Staff across eight of the 34 intervention schools were interviewed. Table 15 provides a breakdown of the achieved sample. Below we discuss the findings in relation to implementation, delivery, perceived outcomes and issues around sustainability.

Table 15: Achieved sample for the process evaluation

School	Teaching assistants	Senior teaching staff	Total staff interviewed for school
1	1	1	2
2	1	1	2
3	1	0	1
4	2	0	2
5	1	0	1
6	1	0	1
7	2	0	2
8	0	1	1
Total	9	3	12

Implementation

This section discusses the implementation of the programme with respect to four factors: (1) the rationale and motivation behind the school's decision to enter the programme, (2) the training received to deliver the programme, (3) the staffing of the programme, and (4) the content of the programme. The section ends by assessing how attractive the intervention is to stakeholders.

School rationale and motivation

There were five key (often interrelated) factors influencing the decision by schools to participate in the programme:

- **Personal interest.** Participation in the programme was led by individual senior teachers' own personal interests in speech and language teaching, which they sought to instil in their schools.
- **Schools' aspirations.** Participation in the programme was positively influenced where schools felt the programme met their schools' objectives and goals. This included seeing the programme as an initiative that placed their school at the centre of advancing knowledge on speech and language teaching and/or seeing the programme as fitting within the wider speech and language strategies at their school.
- **The needs of children.** Schools participated in the programme to meet speech and language needs amongst the children in their school.
- **Whether schools had capacity to deliver the programme.** This related specifically to whether they had staff capacity to deliver the programme. Senior staff sometimes regarded the availability of a suitable teaching assistant who could deliver the programme as a critical factor in their decision to participate.
- **Reputation of the institutions responsible for the programme.** The team that devised the programme came from Higher Education Institutions (HEIs) that were looked upon as highly

reputable by senior school staff. Staff therefore trusted the intentions behind and the quality of product that the HEIs offered and so were willing to take part in the programme.

Training

Training was an important cornerstone of the implementation activities and was attended by all TAs. TAs regarded the training as essential for implementing the intervention. The training programme enabled participants to engage meaningfully with both the manual and programme delivery.

TA training was divided into two events: one event for the nursery intervention, and another for the reception intervention prior to its start. A key difference between these training events was their duration: the nursery training lasted one day while the reception training took place over two days in order to cover the additional components (such as the phonics elements) and sessions (such as the individual sessions).

An overview of the typical format and content of the training as reported by participants is provided below.

- **Training format.** The format of nursery and reception training was similar and involved small groups of TAs (group sizes of 8–12 attendees being reported) from nearby schools working together. The mode of delivery was described by participants as ‘interactive’, with attendees being able to raise queries throughout the event, and being given time to go through the delivery manual without the trainer, either individually or as a group. Practical tasks during the training were limited to observing video demonstrations of programme delivery to children and exercises designed to promote empathy with how children process instructions. Occasionally, participants described a practical exercise that involved the tutor delivering a group session using attendees as stand-in pupils.
- **Training content.** Training typically provided background information about early years’ language learning (for example, what it is, why it is important, and the theory and terminology associated with it) and an overview of the evaluation taking place. This included information about the purpose of the programme and specific guidelines and restrictions governing the trial such as the importance of delivering it according to the manual and not using the programme with non-intervention children.

Participants generally commented favourably on the training, particularly in relation to:

- its duration—participants felt it was the right length (at either a day or two days);
- high-calibre approachable tutors, amenable to answering questions and queries;
- clear and concise delivery of content by tutors;
- the value of small-group training to TA’s from different schools enabling peer-to-peer support; and
- building in time for attendees to engage with the manual in order to consolidate learning.

The findings also indicated that training events could be further refined and developed in order to strengthen their suitability for teaching assistants. Ideas included:

- **Encouraging peer-to-peer support among attendees after training events.** This would provide an additional strand of support for TAs, particularly for those delivering the programme on their own. Our evidence indicated that some participants were already doing this and tended to draw on peer support largely at the start of the programme, as they developed their confidence around delivery. Despite this limited use of peer support¹⁹, it may still be worth tutors actively encouraging and formalising it, for example through ‘buddying’ for

¹⁹ This included both within-school peer support, where schools that had two TAs delivering the programme supported each other, and networking between TAs from different schools that attended the same training event.

teaching assistants that are delivering the programme on their own or encouraging the sharing of attendee details during training events.

- **Introducing more practical elements within the training.** In particular, aspects of training which allow participants more opportunities to observe and deliver the programme, even if just to other attendees. Participants felt that this would have given them valuable insights and experience of delivering the programme prior to beginning it in their schools. However, this suggestion needs to be seen within the context of minimising the time burden on TAs and schools.
- **Reception training.** The findings indicated that the reception training could be modified in the following ways in order to help TAs:
 - training for reception Term 2 should be moved *closer to the delivery* time;
 - equal coverage should be given to the content and delivery of both group and individual sessions; and,
 - for the 30-week intervention, training events should strike a balance between providing enough information for new TAs delivering the programme at reception without being overly-repetitive for those that have already attended the nursery training.

After the training was completed, TAs drew on four types of ongoing support:

- the comprehensive delivery manual;
- support from senior staff within the school to implement the programme;
- observational visits from the project team to ensure that the programme was being correctly implemented; and
- informal peer-to-peer support from TAs delivering the same intervention (as discussed above).

Researchers from the project team also provided support directly to TAs during some of the project team visits.

Generally, schools did not feel they needed much support from the project team after the observational visits. The exception to this was where schools were experiencing significant delivery issues due to both school-related and programme-related factors as discussed in the fidelity section below.

Staffing

Discussions with senior staff provided the following three insights:

- **Appropriate staffing.** Teaching assistants were regarded as the right level of staff to deliver the programme, particularly given the comprehensiveness of the training and the accompanying programme manual. Findings were less clear on whether having one or two TAs in each school worked best: both were seen to have their respective advantages.
- **Motivation as important as experience.** Given the intensive nature of programme delivery, senior teachers identified three key qualities important in selecting suitable TAs: motivation/enthusiasm, reliability, and thoroughness. These were valued over and above previous experience of speech and language teaching.
- **Teaching assistant selection in the school context.** In order to attract TAs within the schools to undertake a role on the programme and to help senior staff make selection decisions, it was seen as important for the programme to speak clearly to the personal development needs of TAs. However, senior staff decisions to appoint TAs also rested on the impact this would have on wider classroom teaching. In particular, even suitably qualified TAs

were not selected to deliver the programme if it was felt that their involvement would disrupt other teaching support within schools.

In order to be able to deliver the programme, TAs required targeted support from senior staff within the school and the programme team. The support of senior staff in protecting TA time, resolving challenges from other teachers' needs, and providing occasional feedback on the delivery of sessions were seen to be important in this regard. Likewise, the programme manual, having a named contact at the programme team to approach in case of emerging issues, and the observational visits made by programme tutors to help refine programme delivery were all seen to be particularly helpful.

Content

Staff involved with the programme identified three particular aspects that they would like to change:

- **Pupil selection.** Some staff had reservations about the types of pupils that were selected for the programme—feeling that some were either too advanced or had learning difficulties that made them unsuitable for the programme. Although acknowledging that pupil selection was based on the needs of the trial, they wanted control over which pupils entered the programme if they continued with it post-trial.
- **Delivery of individual sessions in reception.** These were seen by some staff to be onerous and resource-intensive to deliver to all pupils. It was suggested—if the programme was to continue—that either these should be removed altogether, or that they should be used selectively. This would entail screening, delivering the individual session to only those pupils that, it was felt, needed it.
- **Modifying the resources.** Some of the resources were seen as not being appropriate for use with the selected children. A condition of continuing with the programme was to either to use modified resources from the programme team and/or use school resources, for example those used by speech and language teachers in the school.

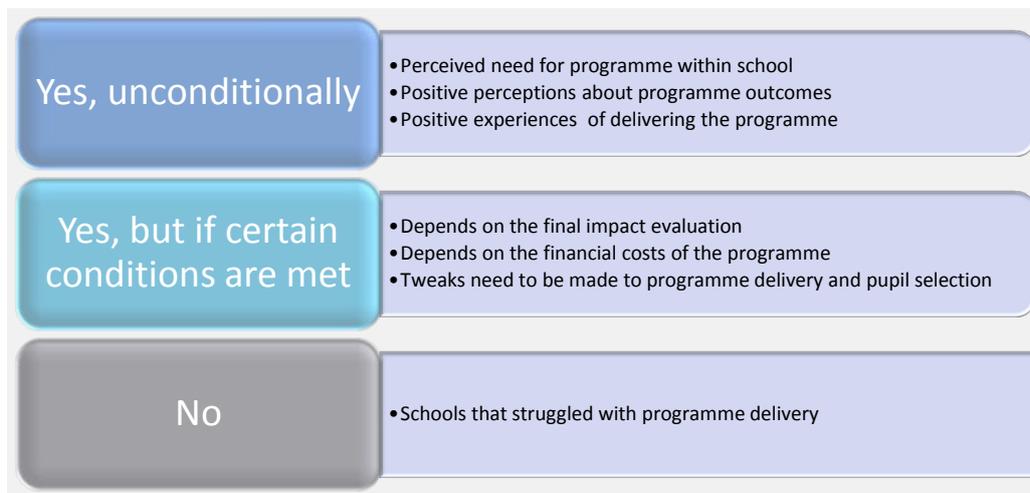
Sustainability

Senior teaching staff and teaching assistants were asked whether they were likely to continue with the programme after the trial. Although the main view was that the programme was a valuable tool in improving the spoken language ability of children, participants' responses hinged on five considerations:

- the perceived need for the programme in the school—including whether other language interventions were already running;
- the perceived or anticipated impact of the programme;
- experiences of delivering the programme—including the use the resources accompanying the programme;
- the eventual financial costs of the programme; and
- the control schools would have in selecting pupils for the programme after the trial.

Accordingly, there were three views on whether schools were willing to continue with programme after the trial. These are organised in order of levels of acceptance in Figure 4 below.

Figure 4: Views on the future of the programme



A key reason for not wanting to carry on with the programme was challenging experiences of delivery. This was particularly felt by participants who struggled to deliver the programme as prescribed and/or found the programme to be too time consuming.

Fidelity

There were varying levels of fidelity to the programme delivery model, where fidelity is defined in terms of adherence to the content and format of delivery as prescribed by the manuals (structure, time, and content). It was clear that some schools would not have been able to continue with the programme had it not been for the flexibility shown by the programme team in allowing certain deviations. Where there were significant fidelity issues, the numbers of sessions delivered varied. For example one to one sessions in reception were omitted in some schools because of resourcing (unfortunately this cannot be quantified across all schools).

There were a number of school- and programme-related factors that affected fidelity and these were tied to the overall delivery of the programme.

Common factors that threatened the fidelity of the programme included:

- **A lack of a discrete space where group and individual sessions could be delivered.** This may be difficult in some schools where space may be limited, but early discussions may ensure that at least some measures are in place.
- **Failure to protect a teaching assistant's time to deliver the programme.** It was clear that teaching assistants had other roles within the school and so the failure to do so had a genuine impact on the fidelity of delivery.
- **A lack of support amongst wider school teaching staff.** This was a challenge for some of the schools, at least initially. Although experienced teaching assistants were able to draw on a number of strategies to deal with this, there is evidence to indicate that the input from the programme team and senior school staff would have been helpful in negotiating this. Suggestions for addressing this included developing information leaflets about the programme and its value to the school for teachers, as well as working with programme coordinators in schools to develop strategies for achieving teacher buy-in.

Factors were also identified that safeguarded against both inadvertent and intentional deviations from the prescribed programme. These included:

- the training events (as described in the training section above) and in particular the strong message concerning the need for adherence;
- a comprehensive delivery manual;
- follow-up visits by the programme team to observe delivery;
- the repetitive framework underpinning the programme; and
- the appropriate selection of pupils (see outcomes section below).

Outcomes

The overwhelming view of school staff was that the programme had a positive impact on pupils. The perceived impacts resonated with the programme's goals of improving spoken language skills and children's confidence. Accordingly, participants reported improvements in: spoken language skills (active listening and conversational ability); vocabulary and general language development; and narrative skills and confidence in terms of children being more outgoing and conversational than prior to the programme.

Participants attributed these observed impacts both to the programme itself, and to factors external to it, such as the ability of pupils selected and the wider teaching taking place within schools. Where the role of the programme was discussed, participants attributed impact to three aspects of the programme: delivery, format and content:

- **Programme delivery.** Positive outcomes were partly attributed to TAs seeing the children regularly and in small groups, as well as individually (although there were more mixed views about the value of individual sessions). This ensured children consistently received support from TAs in a format that was conducive to learning. To a lesser extent, having mixed ability groups was seen to be helpful insofar as more able children were able to support their less able peers.
- **Programme format.** The teaching approach helped to achieve outcomes by making the programme fun and engaging through its use of the listening rules, motivational tools such as the Best Listener Award, and the interactive nature of sessions.
- **Programme content.** The programme content also helped through (a) the use of topics that made sessions less abstract, (b) iteration and consolidation of knowledge across each session, (c) coverage given to a range of components within each session (for example vocabulary and narrative skills), and (d) the overall focus of sessions on narrative and vocabulary work.

Participants were of the view that the programme did not have an equal impact on all pupils. One view was that it was *most effective with pupils who had very low levels of language skills or low confidence* at the start of the programme but *less effective for children with English as an additional language (EAL) or those identified as having learning difficulties (such as autism) or behavioural difficulties*. Although participants were unclear why this was the case, two inferences could be made: first, that having these challenges significantly impacted the delivery of the programme (for example the time needed or the appropriateness of the resources) and so diluted its outcomes, and second, that there were other underlying issues that underpinned language abilities which needed to be addressed separately.

Participants also reported having pupils with more advanced language skills in their groups, which was attributed to the pupil selection process. One view was that the impact on these children was mild and revolved around 'topping-up' or consolidating their existing skills, as opposed to developing new ones. For example, deepening their narrative skills rather than developing them.

Participants were asked to comment on which of the two treatment options (the 30-week or the 20-week), if any, they had a preference for and why. Some staff did not have a view on this and did not

express a preference. Where they did, preferences were based on two interrelated factors: perceptions of what having a shorter/longer programme added to children's experiences and, to a lesser extent, views on the impact of both options on children.

The 30-week programme was preferred on the basis that early exposure to the programme's repetitive framework would lead to familiarity with it by the time children reached reception. They would then be much more settled into the routine. There were two further positive views on the impact of the longer programme on children: (1) the general view that the earlier the children start the programme, the greater the impact it will have on them, and (2) a very tentative view, based on observations, that the 30-week intervention may have had a greater impact. This view was generally based on a 'hunch' that staff had in observing both intervention groups but heavily caveated with the possibility that other factors may have also played a part, such as pupil selection. However, how settled children were by the time they reached reception could be inferred as an 'active ingredient' for this observed difference. Being settled could make children better able to focus on the learning, as opposed to struggling to become familiar with the session structure when they reached reception.

In contrast, staff who preferred the 20-week format viewed the repetitive framework of the 30-week programme as a weakness insofar as children became 'tired' and 'bored' (and hence disengaged) with the sessions by the time they reached reception. From an impact perspective, staff had not observed, or were not convinced, that a longer programme significantly improved the effect that it had on children.

Formative findings

The findings indicate that the following four features of the programme approach and coverage may need to be revisited in order to further aid delivery.

- **Individual sessions.** Delivering both group and individual sessions in reception led to an intense delivery experience for TAs in terms of time spent on the programme and in trying to accommodate these in the school timetable. Although some participants felt that individual sessions should be removed, there was also the view that these could be retained but limited to specific children (for example those that were struggling) rather than the whole group.
- **Topics.** There was a view that some of the topic areas at nursery were too abstract and grown-up for the children and so may need to be modified. Some topic areas were also felt to have sessions that were more interactive and fun than others: this may need to be reviewed to ensure consistency across topic areas.
- **Resources.** Although the resources were generally seen to be of a high quality, some of these may need to be modified to ensure that they are fit for purpose. This includes ensuring: (a) that they reflect the ability and level of children, particularly the nursery resources, (b) that interactive and fun resources are used consistently across the programme, and (c) that they are checked for any implicit socio-economic and cultural biases which may alienate a child.
- **Pupil selection.** Pupils were screened for ability and randomly allocated in this study to one of the intervention options. However, some of the children were seen to be more advanced than others and those with English as a second language and mild learning difficulties also found themselves in the programme. Any future roll-out of the programme needs to consider whether this is desirable and, if not, how best to screen and select pupils.

Conclusion

Key conclusions

1. The Nuffield Early Language Intervention had a positive impact on the language skills of children in the trial. This is true for both the more expensive, 30-week version, starting in nursery, and the 20-week version, delivered only in school.
2. Children receiving the 30-week version experienced the equivalent of about four months of additional progress, compared with about 2 months additional progress for the 20-week version. Both results are unlikely to have occurred by chance, though results for the 30-week version are more secure.
3. The evaluation did not provide reliable evidence that either version of the programme had a positive impact on children's word-level literacy skills.
4. Teaching assistants delivering the programme reported that they found it difficult to devote enough time to it, and that support from senior staff was required to protect the programme time.
5. Staff in participating schools reported that the programme had a positive impact on children's language skills and confidence. They thought that the factors which contributed to this included the small-group format, the activities covered, and the focus on narrative and vocabulary work.

Interpretation

This evaluation found evidence to suggest that both versions of the Nuffield Early Language Intervention had a positive impact on the language skills of pupils with relatively poor language abilities. The effects on language skills were also found to be larger for the 30-week as compared with the shorter 20-week version. By contrast, there is no evidence that the interventions had significant impacts on word-level literacy skills, which is consistent with an earlier randomised controlled trial of the 30-week intervention.²⁰ Both group and individual sessions were usually delivered during lesson time, which entailed pupils being taken out of another class. We should therefore interpret the effects of the interventions relative to a business-as-usual scenario where pupils would have spent the time in the classes.

Interestingly, the estimated effects seem to increase at a follow-up test six months after the end of the trial. This is all the more surprising as some control pupils had already begun to receive an alternative intervention by this point. This suggests that the impacts of the interventions are more likely to grow over time, rather than fade out, with the caveat that we don't know whether any continuing support was being given to the treatment groups during this follow up period.

Qualitative evidence provided by the process evaluation is in line with the empirical results of the impact evaluation. The process evaluation also suggests that effective delivery of the programme depended on staff having adequate time to deliver the programme properly, and on schools having a separate space where group and individual sessions could be delivered. It is notable that staff reported finding it hard to deal with the additional workload caused by delivering the treatments as this suggests the provision of the treatment may have impacted the quality of other learning support provided by TAs. This could have caused negative spillovers on control pupils, but it is hard to judge how large any such spillovers might have been.

Limitations

The main limitations of this evaluation are relatively small sample sizes, attrition, potential spillover effects, limitations to external validity, and uncertainty surrounding the effect sizes at follow-up.

²⁰ Fricke *et al.* (2013).

Considering the main evaluation sample of 350 pupils, the actual minimum detectable effect size was fairly high at 0.25. However, as the estimated treatment effect exceeded this threshold, the impact evaluation was able to detect significant effects despite the small scale of the intervention. Nevertheless, the confidence intervals are relatively wide, meaning that there is considerable uncertainty as to the precise effect of the intervention (the 95% confidence interval for the effect size of the 30-week intervention varies from 0.07 to 0.46).

Small sample sizes imposed even greater limitations on the subgroup analyses. In particular, restricting analysis to FSM eligible pupils resulted in a sample of 66 pupils. Such a small sample size both considerably increases the minimum detectable effect size and reduces the ability for randomisation to generate well-balanced groups. Indeed, the treatment and control group were heavily imbalanced among FSM eligible pupils. Future EEF trials should consider randomisation strategies that ensure sub-groups are balanced if sub-group analysis is an explicit goal of the intervention, particularly in the case of small-scale trials.

A further limitation was the attrition of both pupils and schools. Pupils who changed schools between nursery and primary school dropped out of the trial and could not be followed up at the post-test stage. Three schools also withdrew from the trial, citing significant staffing issues created by the intervention. Pupils in these schools could be followed up and included in the analysis as post-test data was still recorded for these pupils. However, it does mean that our impact estimates include the effects of only partial programme delivery for three schools, though the results are largely unchanged if we exclude the schools that dropped out.

Because randomisation was conducted within-school at the pupil level, a further potential limitation was the risk of spillover effects whereby pupils in the control group were also impacted in some way by the treatment. Such spillover effects could either be positive (for example if TAs applied the treatment teaching methods outside of the targeted programme sessions), or negative (for example if TAs were less able to support control group pupils in 'business as usual' because of the time pressures that delivering the programme placed on them). Positive spillover effects would cause treatment effects to *underestimate* the impact of the programme, whereas negative treatment effects would cause treatment effects to *overestimate* the impact of the programme. The process evaluation reported that TAs appreciated the importance of confining the treatment to the programme sessions, reducing the likelihood of positive spillovers, although it also emphasised that TAs struggled with the extra time commitment the treatment required, which suggests negative spillover effects may limit the results presented here.

The experimental context is also important when considering external validity. The average pupil to TA ratio among schools involved in the trial was significantly lower than the national average for primary schools in England. This suggests that the delivery problems caused by increased TA workloads, as highlighted in the process evaluation, may be more severe if the programmes are implemented in schools that are closer to the national average.

Finally, there is also some uncertainty as to what mechanisms are driving the increased treatment effects at the follow-up stage. This is because of the large number of differences in experimental conditions at follow-up stage: some schools had already begun to deliver an alternative intervention, the interventions were delivered at different stages and ages, and it is not clear whether TAs continued to use the intervention materials and techniques with treatment pupils.

Future research and publications

The findings of the Nuffield Early Language Intervention impact evaluation indicate several areas that could benefit from further research. The lack of any significant estimated effect on word-level literacy suggests that improving language skills does not necessarily improve literacy skills for children transitioning from nursery to reception, at least in the short run. If improving the literacy skills of low-

achieving pupils is regarded as a priority within primary schools, it would be useful to conduct future interventions that explicitly target literacy. Furthermore, it will be useful to follow these pupils over time to examine any lasting impacts at Key Stage 2 (and hopefully more schools will provide the necessary data to enable such analysis).

While the relationship between pupil-teacher ratios and pupil attainment has been examined in some depth,²¹ evidence on the impact of non-teaching staff members on pupil attainment has been relatively scarce until recently. However, with growing numbers of teaching assistants within schools, there is a need to find the most effective ways to deploy them. This evaluation suggests that providing intense language support sessions to pupils with low language skills is one such method. This adds to a growing body of evidence suggesting that the delivery of structured interventions might be a relatively effective way to use TAs.²² By testing other programmes that are delivered by non-teaching staff, future interventions could make valuable contributions.

Small sample sizes meant that this evaluation was unable to find secure treatment effect estimates for FSM eligible pupils. As this subsample of relatively deprived pupils is of particular interest to the EEF, these findings emphasise that future interventions should aim for larger sample sizes, or consider including FSM eligibility in the randomisation process to enable robust subgroup analysis.

Lastly, we observe that the difference between the language skills of treatment and control school pupils grew between the initial post-test and the follow-up test conducted after six months. This is consistent with the effect size increasing over time, rather than fading out. As a result, for future similar trials, it might be worth considering conducting post-tests with a slight delay rather than at the end of the intervention, but before any alternative interventions have been introduced. This might increase the chance of finding (true) larger effects.

²¹ For an overview see DfE (2011).

²² For example, see the latest EEF guidance on 'Making Best Use of Teaching Assistants' (<https://educationendowmentfoundation.org.uk/news/teaching-assistants-should-not-be-substitute-teachers-but-can-make-a-real-d/>)

References

- Altman, D. G. and Bland, J. M. (2005) 'Treatment allocation by minimisation'. *BMJ: British Medical Journal*, 330: 7495, 843.
- Blundell, R., Dearden, L. and Sianesi, B. (2005) 'Evaluating the effects of education on earnings: models, methods and results from the National Child Development Survey'. *Journal of the Royal Statistical Society Series A*, Royal Statistical Society, 168: 3, 473–512.
- Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., Götz, K. and Hulme, C. (2008) 'Improving early language and literacy skills: differential effects of an oral language versus a phonology with reading intervention'. *Journal of Child Psychology and Psychiatry*, 49: 4, 422–432.
- Bruhn, M. and McKenzie, D. (2009) 'In Pursuit of Balance: Randomization in Practice in Development Field Experiments'. *American Economic Journal: Applied Economics*, 1: 4, 200–232.
- Clarke, P. J., Snowling, M. J., Truelove, E. and Hulme, C. (2010) 'Ameliorating Children's Reading-Comprehension Difficulties: A Randomized Controlled Trial'. *Psychological Science*, 21: 8, 1106–1116.
- DfE: Department for Education (2011) 'Class size and education in England: evidence report'. Research report DFE-RR169. Available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/183364/DFE-RR169.pdf
- Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C. and Snowling, M. J. (2013) 'Efficacy of language intervention in the early years'. *The Journal of Child Psychology and Psychiatry*, 54: 3, 280–290.
- Hulme, C. and Snowling, M. J. (2013) 'Learning to Read: What We Know and What We Need to Understand Better'. *Child Development Perspectives*, 7: 1, 1–5.
- Jenkins, A., Levačić, R. and Vignoles, A. (2006) 'Estimating the Relationship between School Resources and Pupil Attainment at GCSE'. Institute of Education, University of London, DfES Research Report RR727. Available at: <http://webarchive.nationalarchives.gov.uk/20130401151715/http://www.education.gov.uk/publications/eOrderingDownload/RR727.pdf>
- Morgan, K.L. and Rubin, D.B. (2012) 'Rerandomization to improve covariate balance in experiments'. *Annals of Statistics*, 40: 2, 1263–1282.
- NDCS: National Deaf Children's Society (2014) 'Assessing and monitoring the progress of deaf children and young people: Communication, language and listening'. London: National Deaf Children's Society. Available at: http://www.ndcs.org.uk/professional_support/our_resources/assessments.html
- Nicoletti, C. and Rabe, B. (2013), 'Productivity of school expenditure: Differences across pupils from diverse backgrounds'. Institute for Economic and Social Research. Available at: http://www.iwae.org/papers%20sito%202013/Rabe_c.pdf
- Paslowski, T. (2005) 'The Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4): A Review'. *Canadian Journal of School Psychology*, 20: 1–2, 129–134.
- Raudenbush, S. W. and Bryk, A. S. (2002) *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.
- Renfrew, C. (2010) *Action Picture Test Revised Edition*. Buckingham: Hinton House Publishers Ltd.
- Saghaei, M. (2011) 'An Overview of Randomization and Minimization Programs for Randomized Clinical Trials'. *Journal of Medical Signals and Sensors*, 1: 1, 55–61.

Semel, E., Wiig, E. H. and Secord, W. A. (2003) *Clinical evaluation of language fundamentals, fourth edition (CELF-4)*. Toronto: The Psychological Corporation/A Harcourt Assessment Company.

Scarborough, Hollis (2009) 'Connecting Early Language and Literacy to Later Reading (Dis)abilities: Evidence, Theory, and Practice' in Fletcher-Campbell, F., Soler, J. and Reid, G. (eds) *Approaching Difficulties in Literacy Development: Assessment, Pedagogy and Programmes*. London: Sage.

Scott, N. W., McPherson, G. C., Ramsay, C. R. and Campbell, M. K. (2002) 'The method of minimization for allocation to clinical trials. a review'. *Control Clinical Trials*, 23, 662–674.

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove E., Nation, K. and Hulme, C. (2009) *YARC: York Assessment of Reading for Comprehension Passage Reading*. Swindon: GL Publishers.

Snowling, M. J., and Hulme C. (2011) 'Evidence-based interventions for reading and language difficulties: Creating a virtuous circle'. *British Journal of Educational Psychology*, 81, 1–23.

Whitehurst, G. J. and Lonigan, C. J. (1998) 'Child Development and Emergent Literacy'. *Child Development* 69: 3 (June 1998), 848–872.

Wooldridge, J. (2010) *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). MIT Press.

Appendix A: Additional tables

Table A1: Regression of control variables on primary language composite outcome amongst control pupils

	(1) Language Composite	(2) Language Composite
Age in months	0.0597**	-0.00322
	[0.0261]	[0.0169]
Female	0.284	0.00898
	[0.180]	[0.116]
English as an Additional Language	0.0103	0.147
	[0.227]	[0.147]
Known Speech and Language Difficulties	1.112***	0.958***
	[0.268]	[0.170]
Language Composite (std)		-
CELF expressive vocabulary (std)		0.0768
		[0.119]
APT information (std)		0.152
		[0.109]
APT grammar (std)		0.220**
		[0.0949]
Listening comprehension (std)		0.191**
		[0.0759]
CELF sentence structure (std)		-0.0199
		[0.0845]
Nursery Vocabulary Naming (std)		0.0791
		[0.110]
British Picture Vocabulary Scale (std)		0.204*
		[0.106]
Nursery Vocabulary Definitions (std)		0.0535
		[0.0951]
Constant	-3.098**	-0.049
Observations	115	115
R-Squared	0.101	0.552

Note: * indicates that the difference in means is significant at the 10% level ** at the 5% level *** at the 1% level. Standard deviations are reported in square brackets.

Appendix B: Subsample balance analysis

Table B1: Pupils without known speech and language difficulties or EAL
Average baseline characteristics of groups (differences calculated as effect sizes)

Characteristic	All	30-week treatment (T1)	20-week treatment (T2)	Control	Diff T1 – C	Diff T2 – C
<u>Characteristics considered at randomisation</u>						
Proportion of pupils who are female	0.48	0.47	0.48	0.49	-0.04	-0.02
	[0.50]	[0.50]	[0.50]	[0.50]		
Age in months	45.99	45.95	45.77	46.26	-0.09	-0.14
	[3.51]	[3.38]	[3.56]	[3.62]		
Screening language composite	6.7	6.82	6.77	6.51	0.17	0.14
	[1.88]	[1.79]	[1.91]	[1.93]		
	[0.00]	[0.00]	[0.00]	[0.00]		
<u>Primary composite baseline observation and components</u>						
Language composite (primary outcome)	0.06	0.12	0.06	0	0.12	0.06
	[0.98]	[0.94]	[1.07]	[0.93]		
CELF expressive vocabulary	11.29	11.49	11.64	10.75	0.14	0.17
	[5.21]	[4.97]	[5.60]	[5.06]		
APT information	20.7	21.16	20.34	20.6	0.09	-0.04
	[5.97]	[6.01]	[5.88]	[6.06]		
APT grammar	14.04	14.81	13.5	13.81	0.17	-0.05
	[5.85]	[6.09]	[5.77]	[5.68]		
Listening comprehension	1.37	1.28	1.5	1.32	-0.03	0.11
	[1.63]	[1.61]	[1.93]	[1.32]		
<u>Other baseline scores</u>						
CELF sentence structure	6.72	7.02	6.5	6.65	0.1	-0.04
	[3.78]	[3.81]	[3.69]	[3.86]		
Vocabulary naming	5.8	5.75	5.99	5.68	0.03	0.13
	[2.29]	[2.29]	[2.28]	[2.30]		
Vocabulary definitions	4.48	4.41	4.44	4.6	-0.06	-0.05
	[3.28]	[3.43]	[3.26]	[3.16]		
British picture vocabulary scale	37.94	37.93	39.15	36.78	0.08	0.17
	[13.97]	[12.06]	[14.57]	[15.15]		
n	286	96	94	96		

Note: * indicates that the difference in means is significant at the 10% level ** at the 5% level *** at the 1% level. Standard deviations are reported in square brackets.

Table B2: Male pupils
Average baseline characteristics of groups (differences calculated as effect sizes)

Characteristic	All	30-week treatment (T1)	20-week treatment (T2)	Control	Diff T1 – C	Diff T2 – C
<i>Characteristics considered at randomisation</i>						
Age in months	46.32	46.59	46.32	46.05	0.15	0.07
	[3.63]	[3.50]	[3.75]	[3.67]		
Screening language composite	6.43	6.6	6.4	6.28	0.17	0.06
	[1.96]	[1.82]	[2.04]	[2.02]		
<i>Other pupil characteristics</i>						
Proportion of pupils with EAL	0.13	0.09	0.17	0.14	-0.14	0.11
	[0.34]	[0.28]	[0.38]	[0.35]		
Proportion of pupils with known speech and language difficulties	0.04	0.03	0.05	0.03	0	0.07
	[0.19]	[0.18]	[0.21]	[0.18]		
<i>Primary composite baseline observation and components</i>						
Language composite (primary outcome)	-0.1	0.03	-0.1	-0.23	0.26	0.13
	[1.06]	[1.02]	[1.22]	[0.90]		
CELF expressive vocabulary	10.59	11.57	10.56	9.66	0.36**	0.17
	[5.38]	[4.96]	[5.99]	[5.01]		
APT information	19.8	20.41	19.68	19.32	0.18	0.06
	[6.56]	[7.20]	[6.82]	[5.63]		
APT grammar	13.58	14.45	13.1	13.23	0.21	-0.02
	[6.34]	[6.80]	[6.60]	[5.57]		
Listening comprehension	1.17	1.12	1.35	1.03	0.05	0.2
	[1.61]	[1.52]	[1.96]	[1.23]		
<i>Other baseline scores</i>						
CELF sentence structure	6.61	6.52	6.57	6.75	-0.06	-0.05
	[3.80]	[3.71]	[3.78]	[3.95]		
Vocabulary naming	5.36	5.74	5.21	5.14	0.26	0.03
	[2.45]	[2.52]	[2.50]	[2.32]		
Vocabulary definitions	4.17	4.49	4.07	3.95	0.16	0.04
	[3.39]	[3.77]	[3.43]	[2.96]		
British picture vocabulary scale	36.41	36.55	37.71	34.86	0.12	0.2
	[14.83]	[13.43]	[15.84]	[15.12]		
n	180	58	63	59		

*Note: * indicates that the difference in means is significant at the 10% level ** at the 5% level *** at the 1% level. Standard deviations are reported in square brackets.*

Table B3: Female pupils
Average baseline characteristics of groups (differences calculated as effect sizes)

Characteristic	All	30-week treatment (T1)	20-week treatment (T2)	Control	Diff T1 – C	Diff T2 – C
<i>Characteristics considered at randomisation</i>						
Age in months	45.86	45.3	45.86	46.43	-0.33*	-0.16
	[3.43]	[3.19]	[3.46]	[3.61]		
Screening language composite	6.62	6.63	6.7	6.53	0.05	0.09
	[1.92]	[1.79]	[2.02]	[1.96]		
<i>Other pupil characteristics</i>						
Proportion of pupils with EAL	0.17	0.18	0.19	0.14	0.09	0.12
	[0.38]	[0.39]	[0.40]	[0.35]		
Proportion of pupils with known speech and language difficulties	0.02	0.02	0.03	0.02	0	0.11
	[0.15]	[0.13]	[0.18]	[0.13]		
<i>Primary composite baseline observation and components</i>						
Language composite (primary outcome)	0.1	0	0.11	0.2	-0.2	-0.09
	[0.92]	[0.91]	[0.94]	[0.93]		
CELF expressive vocabulary	11.1	10.32	11.26	11.71	-0.26	-0.09
	[5.13]	[5.01]	[5.20]	[5.18]		
APT information	21.06	20.89	20.73	21.55	-0.11	-0.13
	[5.54]	[5.08]	[5.28]	[6.26]		
APT grammar	14.26	14.31	14.06	14.4	-0.02	-0.06
	[5.36]	[5.45]	[5.26]	[5.48]		
Listening comprehension	1.51	1.25	1.64	1.64	-0.25	0
	[1.58]	[1.56]	[1.70]	[1.46]		
<i>Other baseline scores</i>						
CELF sentence structure	6.5	6.93	6.53	6.04	0.24	0.13
	[3.76]	[3.74]	[3.90]	[3.65]		
Vocabulary naming	5.72	5.25	6.05	5.84	-0.25	0.09
	[2.21]	[2.27]	[2.02]	[2.29]		
Vocabulary definitions	4.69	3.87	5.06	5.14	-0.37**	-0.02
	[3.46]	[3.21]	[3.81]	[3.21]		
British picture vocabulary scale	38.01	36.88	39.03	38.07	-0.09	0.07
	[12.87]	[11.83]	[12.44]	[14.36]		
n	170	56	58	56		
<i>Note: * indicates that the difference in means is significant at the 10% level ** at the 5% level *** at the 1% level. Standard deviations are reported in square brackets.</i>						

Table B4: Pupils eligible for FSM in reception
Average baseline characteristics of groups (differences calculated as effect sizes)

Characteristic	All	30-week treatment (T1)	20-week treatment (T2)	Control	Diff T1 – C	Diff T2 – C
Characteristics considered at randomisation						
Proportion of pupils who are female	0.5	0.3	0.54	0.67	-0.73	-0.26
	[0.50]	[0.47]	[0.51]	[0.49]		
Age in months	45.76	46.4	45.79	45	0.42	0.24
	[3.32]	[2.66]	[3.47]	[3.74]		
Screening language composite	5.98	6.03	5.98	5.94	0.04	0.02
	[1.84]	[1.78]	[1.86]	[1.97]		
Proportion of pupils with EAL	0.12	0.05	0.21	0.06	-0.02	0.48
	[0.33]	[0.22]	[0.42]	[0.24]		
Proportion of pupils with known speech and language difficulties	0.05	0.05	0.07	0	0.24	0.34
	[0.21]	[0.22]	[0.26]	[0.00]		
Primary composite baseline observation and components						
Language composite (primary outcome)	-0.27	-0.21	-0.46	-0.06	-0.15	-0.4
	[0.96]	[1.08]	[0.82]	[1.02]		
CELF expressive vocabulary	9.83	10.45	8.93	10.56	-0.02	-0.31
	[5.02]	[4.71]	[5.42]	[4.76]		
APT information	19.25	18.5	19.07	20.36	-0.3	-0.21
	[6.42]	[7.38]	[5.85]	[6.34]		
APT grammar	13.05	13.78	12.04	13.81	-0.01	-0.3
	[6.57]	[7.88]	[5.36]	[6.86]		
Listening comprehension	0.8	0.95	0.46	1.17	-0.14	-0.44*
	[1.21]	[1.36]	[0.79]	[1.47]		
Other baseline scores						
CELF sentence structure	5.47	5.15	6.46	4.28	0.23	0.58**
	[3.43]	[2.81]	[3.66]	[3.41]		
Vocabulary naming	5.21	5.45	4.96	5.33	0.05	-0.16
	[2.17]	[2.48]	[1.57]	[2.66]		
Vocabulary definitions	3.64	3.38	3.79	3.69	-0.09	0.03
	[3.46]	[2.68]	[3.75]	[3.90]		
British picture vocabulary scale	32.74	34.2	33.04	30.67	0.25	0.17
	[12.14]	[10.53]	[10.45]	[16.10]		
n	66	20	28	18		

Note: * indicates that the difference in means is significant at the 10% level ** at the 5% level *** at the 1% level. Standard deviations are reported in square brackets.

Table B5: Average baseline NPD characteristics of groups – FSM eligible pupils only (differences calculated as effect sizes)

Characteristic	All	30-week treatment (T1)	20-week treatment (T2)	Control	Diff T1 – C	Diff T2 – C
Proportion of pupils recorded as SEN action/plus	0.167	0.35	0.036	0.167	0.49	-0.35
	[0.376]	[0.489]	[0.189]	[0.383]		
Proportion of pupils of White British ethnicity	0.773	0.85	0.607	0.944	-0.22	-0.8
	[0.422]	[0.366]	[0.497]	[0.236]		
n	66	20	28	18		
<i>Note: * indicates that the difference in means is significant at the 10% level ** at the 5% level *** at the 1% level. Standard deviations are reported in square brackets.</i>						

Appendix C: Alternative methodologies

Table C1: Alternative treatment effect estimates for all pupils and across sub-groups for Language composite (primary outcome)

	Methodology					
	(1) Raw	(2) OLS	(3) Random Effects	(4) Fixed Effects	(5) FILM	(6) Kernel Matching
<u>20 week treatment</u>						
All pupils	0.280**	0.266***	0.248***	0.235**	0.267***	0.275**
	[0.133]	[0.096]	[0.096]	[0.096]	[0.099]	[0.126]
Pupils without known speech and language difficulties or EAL	0.365***	0.322***	0.299***	0.296***	0.301***	0.322**
	[0.133]	[0.094]	[0.090]	[0.093]	[0.095]	[0.140]
Male pupils	0.308	0.14	0.117	0.047	0.117	0.125
	[0.199]	[0.135]	[0.127]	[0.123]	[0.134]	[0.175]
Female pupils	0.248	0.354**	0.329*	0.206	0.374**	0.291
	[0.207]	[0.163]	[0.168]	[0.178]	[0.164]	[0.232]
Pupils with observed NPD records	0.2	0.248*	0.269**	0.276**	0.266**	0.321*
	[0.176]	[0.120]	[0.124]	[0.125]	[0.127]	[0.180]
Including NPD covariates	0.2	0.248*	0.254*	0.260*	0.295**	0.277
	[0.176]	[0.130]	[0.132]	[0.130]	[0.130]	[0.197]
<u>30 week treatment</u>						
All pupils	0.222**	0.176*	0.176**	0.188**	0.161*	0.184
	[0.104]	[0.091]	[0.090]	[0.090]	[0.091]	[0.118]
Pupils without known speech and language difficulties or EAL	0.325***	0.273**	0.251**	0.225**	0.260**	0.239*
	[0.109]	[0.112]	[0.101]	[0.097]	[0.118]	[0.139]
Male pupils	0.310**	0.229*	0.224*	0.300**	0.202*	0.262
	[0.127]	[0.122]	[0.124]	[0.127]	[0.113]	[0.181]
Female pupils	0.132	0.121	0.114	0.065	0.102	0.087
	[0.195]	[0.132]	[0.138]	[0.156]	[0.125]	[0.195]
Pupils with observed NPD records	0.214	0.152	0.139	0.212*	0.167	0.133
	[0.128]	[0.127]	[0.130]	[0.124]	[0.141]	[0.196]
Including NPD covariates	0.214	0.171	0.17	0.224	0.169	0.132
	[0.128]	[0.129]	[0.136]	[0.132]	[0.155]	[0.203]
<p>Note: * indicates that the difference in means is significant at the 10% level ** at the 5% level *** at the 1% level. Standard errors are clustered at the school level and are reported in square brackets. All outcomes are standardised within the estimation sample prior to estimation. The following covariates are used for all pupils in columns (2)-(6): age in months, gender, whether pupils have English as an additional language, whether they have known speech or language difficulties, each component of the pre-test language composite outcome and each component of the secondary outcome. Covariates from the National Pupil Database include eligibility for Free School Meals, major ethnic group, whether they have ever had special educational needs and IDACI percentile. Column (2) controls for covariates using Ordinary Least Squares. Column (3) allows for a school effect that is uncorrelated with covariates. Column (4) estimates a fixed effect for each school. Column (5) allows the treatment effect to linearly interact with the treatment. Column (6) uses kernel propensity score matching to balance the samples.</p>						

Appendix D: School and parent consent letters



THE UNIVERSITY *of York*



Parent/Carer Consent Form

The Nuffield Early Language Intervention

Please initial the boxes below if you are happy for your child to take part in the project *The Nuffield Early Language Intervention*:

Please initial

box

I have read and understood the information given about the Nuffield Early Language Intervention project, and have had the opportunity to ask questions about the project.

I understand that I will be informed if my child is selected to take part in the project.

I understand that if selected my child will receive language intervention in nursery and Reception OR

In Reception only OR language and reading intervention in Year 1.

I understand that my child will be regularly assessed, both individually and in groups, over the course of the project (April 2013 to December 2014).

I understand that the research team may record assessment sessions for analysis purposes only. I understand that these audio recordings will be kept confidential and that no person outside the research team will have access to them.

I understand that my child will be withdrawn from the classroom to receive this intervention.

I agree that the research team can collect the EYFS Profile score from my child's school records.

I understand that participation is voluntary and I reserve the right to withdraw my child at any stage.

I understand that such information will be treated as strictly confidential and handled in accordance

with the provisions of the Data Protection Act 1998. I understand that the information gained will be anonymous and that children's names will be removed from any materials used in the research. I consent to the processing of my child's personal information for the purposes of this research study.

I agree for my child _____ to take part in this study.
(child's name)



_____	_____	_____
Date	Signature (Parent/Carer)	Print Name (Parent/Carer)

_____	_____	_____
Date	Signature (Researcher)	Print Name (Researcher)

The University of Sheffield
Department of Human Communication Sciences
362 Mushroom Lane
Sheffield
S10 2TS

Dr Silke Fricke
PhD, MSc, SLT

Telephone: +44 (0) 114 222 2419
Email: S.Fricke@sheffield.ac.uk

NAME OF HEADTEACHER
SCHOOL
STREET
CITY
POSTCODE

Date

Dear HEADTEACHER

Project: Nuffield Early Language Intervention

Thank you for your school's support in carrying out the *Nuffield Early Language Intervention* research project. As the Reception intervention gets underway this term, we hope that you, your staff and pupils are looking forward to your continued involvement in this project.

As you know, this is an exciting project, evaluating a new intervention approach to supporting language development. The project is funded by the Education Endowment Foundation (EEF), a charity which aims to support children's educational achievement. In addition to the work we are doing to understand how much children's language and literacy skills improve as a result of the intervention, a separate evaluation team are providing independent validation of our results. This will offer more detailed insights into which children the intervention works best for and which elements of the intervention are most successful. The independent evaluation team for this intervention comprises researchers from the Institute for Fiscal Studies (IFS) and NatCen Social Research, and we would like to ask for your assistance in working with these two organisations.

The evaluation team would like to use some of the information about children held in the National Pupil Database (NPD) in order to inform their assessment of the effectiveness of the intervention. NPD records are held by the Department for Education and comprise information that you provide each term as part of the census return – such as ethnicity and eligibility for free school meals – plus children's Key Stage test results. This information will form an integral part of the evaluation and will also allow EEF to continue to follow the progress of children involved in the study after it has ended. In order for the evaluation team to access these records, we need your permission. If you are happy for us to use this data, then we would be grateful if you could provide us with the Unique Pupil Number, name and postcode of all children involved in the study, which we can then pass on to IFS in order for them to identify the correct children in the National Pupil Database.

We would like to reassure you that all data will be treated with the strictest confidence and that no individual pupil or school will ever be identified in any report arising from the research. We should also highlight that the matching of pupil information to NPD records would be carried out by a specialist team at the Department for Education, who would return anonymous pupil records to IFS and EEF, meaning that they would no longer be able to identify pupils by name.

In addition, researchers from NatCen will need to carry out a number of interviews with the practitioners involved in the study to explore views and experiences of delivering the programme. This is essential for us to understand which parts of the intervention worked particularly well and which could be improved. It also offers you a valuable opportunity to provide your views of the programme and its success. NatCen will select eight case study schools to take part in these interviews.

- In the first four schools, they would like to carry out face-to-face interviews with teaching assistants (lasting approximately 1 hour) and with classroom teachers or Foundation Stage co-ordinators (lasting 30 minutes).
- In the remaining four schools, they will carry out telephone interviews with the teaching assistants delivering the programmes. These interviews will last approximately 45-60 minutes.

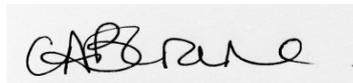
All interviews will be carried out in the spring term of 2014. We are writing to seek your consent for your school to participate in this part of the project. If you agree, we will inform NatCen who will select the schools to take part. If your school is selected, they will contact you directly to make arrangements to carry out the interviews at a time to suit you. All information will be treated as confidential. Again, we would like to reassure you that neither individual staff members nor schools will be identifiable from the data obtained through these interviews.

We are very grateful for your assistance in evaluating the *Nuffield Early Language Intervention*, as it will help us to improve the intervention for other pupils in future. Please sign and date the form below indicating your willingness to participate in these two parts of the project and return it to Silke Fricke at the address at the top of this letter as soon as possible.

Thank you for taking the time to read through the information given above. If you have any questions please do not hesitate to contact me at claudine.bowyer-crane@york.ac.uk, telephone 01904 324398. We look forward to continuing our work with you on this project.

Kind regards


Dr Silke Fricke
The University of Sheffield


Dr Claudine Bowyer-Crane
University of York

HEAD TEACHER CONSENT FORM
PROJECT: NUFFIELD EARLY LANGUAGE INTERVENTION
Independent evaluation

Please complete and sign this form and return it to Silke Fricke (Department of Human Communication Sciences, The University of Sheffield, 362 Mushroom Lane, Sheffield S10 2TS).

Name of school: NAME OF SCHOOL

Name of Headteacher _____

Researchers from UCL London and the University of Sheffield are conducting a project evaluating the effectiveness of The Nuffield Early Language Intervention programme. The intervention is being independently evaluated by the Institute for Fiscal Studies and NatCen Social Research. I have read and understood the information given to me about the independent evaluation and give my permission for NAME OF SCHOOL to take part as indicated below (please tick the appropriate boxes).

I agree to provide the research team with the Unique Pupil Number, name and postcode of all children involved in the study which they can pass on to IFS in order to access the National Pupil Database.

Yes No

I agree to take part in the interviews carried out by NatCen if Name of School is selected and am happy for NatCen to contact us to arrange these interviews.

Yes No

I have been informed about the aims and procedures involved in this research. I reserve the right to withdraw any child at any stage in the proceedings, and also to terminate the project altogether if I think it is necessary. I understand that the information used in the evaluation will be anonymous and that the names of the children, staff members and school will not appear in any outputs from this research.

 Date Signature (Head teacher) Print Name (Head teacher)

The University of Sheffield
Department of Human Communication Sciences
31 Claremont Crescent
Sheffield S10 2TA UK

Dr Silke Fricke
PhD, MSc, SLT

Telephone: +44 (0) 114 222 2419
Email: S.Fricke@sheffield.ac.uk

Name
Head teacher or Deputy Head
School
Street
City
Postcode

Date

Dear Name

Project: Nuffield Early Language Intervention

Thank you for agreeing to take part in the *Nuffield Early Language Intervention* research project. We hope that you, your staff and pupils will enjoy being involved. As you might remember I CAN have partnered with Professor Maggie Snowling and Professor Charles Hulme, and their research teams to conduct the project. The research team coordinating the project in the north of England is based at the University of Sheffield (Department of Human Communication Sciences; 31 Claremont Crescent; Sheffield S10 2TA) and is managed by two people:

Dr Silke Fricke (S.Fricke@sheffield.ac.uk; 0114 – 2222419)

Dr Claudine Bowyer-Crane (claudine.bowyer-crane@york.ac.uk; 01904 – 434398)

We have now received the project agreement for your school from I CAN and we look forward to working with you on this valuable project which we hope will be of benefit to many children. While I CAN will provide the training in delivering the Nuffield Early Language Intervention to all teaching assistants selected by participating schools, the research team is responsible for carrying out the assessments and collecting the data in the participating schools. Details of the anticipated timeline of assessments are contained in this letter. We hope that all the information provided is clear and accessible. However, if you have any queries about any of the information, please do not hesitate to contact us.

Timeline

It is anticipated that the project will begin in January 2013. The information below indicates when the research team would like to visit your school to assess the children.

- In the initial screening phase (Jan – Feb 2013), we would like to see all children who are due to start Reception in the next academic year (i.e. Sept 2013). Children will be seen individually for approximately 10-15 minutes per child.
- We will select 12 children in each school based on their performance on the language screening measures.

- We anticipate that children selected as suitable for the project will be assessed at three separate time points in addition to the initial screening phase (i.e. pre-test Feb - Mar 2013, post-test Apr - Mar 2014, and follow-up test Oct - Dec 2014).
- Each test phase will involve seeing all 12 children on an individual basis and may also involve some group testing. Testing will be carried out with due consideration of the children's attention span.
- Assessments will be carried out by members of the Research Team from the University of Sheffield, the University of York or University College London. All members of the team will be fully trained and will have enhanced CRB clearance before entering the school.
- We will also carry out assessments with children in the year above those participating in the intervention to act as a control cohort. These children will be seen twice over the course of the project (Mar – May 2013; Apr – July 2014)

Thus, the first phase of the project involves screening all children in your nursery who are due to start Reception in the next academic year (i.e. Sept 2013). We will be in contact with you as soon as schools are back after Christmas to start making arrangements for this first phase (screening).

Ethical Approval and Consent

The project has been ethically approved by the research ethics committee of the University College London (UCL). We anticipate carrying out the initial screening with head teacher consent (please see enclosed head teacher consent form).

Written parent/carer consent will be sought for subsequent assessment phases (pre-, post-, and follow-up test) and for participation in the project and the intervention programme. Following the initial screening, children will not be seen at any point unless the research team has first received the parent/carer consent form.

Next steps

We would be grateful if you could complete and sign the Head Teacher Consent Form and return it to Silke Fricke at the Department of Human Communication Sciences as soon as possible. We will then contact you after the Christmas break to make screening-related arrangements with each setting (e.g. finding a day that is convenient for your setting to screen the children).

We are grateful to all schools who are taking part in the study. If you have any questions regarding the project's data collection, please feel free to contact us via phone or email. For questions related to the training of teaching assistants or the delivery of the Nuffield Early Language Intervention, please contact I CAN.

Thank you for taking the time to read through the information given above. We hope that you and your staff will benefit from taking part in the project.

Kind regards



Dr Silke Fricke
The University of Sheffield



Dr Claudine Bowyer-Crane
University of York

HEAD TEACHER CONSENT FORM
PROJECT: NUFFIELD EARLY LANGUAGE INTERVENTION

Please complete and sign this form and return it to Silke Fricke (Department of Human Communication Sciences, The University of Sheffield, 31 Claremont Crescent, Sheffield S10 2TA).

Name of school: _____ Contact person for project (if different from head teacher): _____ _____ Position in setting (contact person): _____ Tel (contact person): _____ Email (contact person): _____ _____

Researchers from UCL London and the University of Sheffield are conducting a project evaluating the effectiveness of The Nuffield Early Language Intervention programme. I have read and understood the information given to me about the project by I CAN and the research team and give my permission for _____ to take part.

I understand that the project will involve the following:

1. Initial screening of all nursery children to identify those suitable to participate in this research
2. Regular assessment of all children taking part in the project and a comparison group of children who are a year older than the children receiving the intervention
3. Delivery of intervention programmes in nursery and Reception classes by a trained teaching assistant who is selected by the school
4. Telephone support for teaching assistants delivering the interventions

Consent for initial screening:

I agree to the initial screening being carried out with head teacher consent. I am aware that parent/carers consent will be sought for subsequent assessment and project phases.

I have been informed about the aims and procedures involved in this research. I reserve the right to withdraw any child at any stage in the proceedings, and also to terminate the project altogether if I think it is necessary. I understand that the information gained will be anonymous and that the children's names and the school's name will be removed from any materials used in this research.

Date	Signature (Head teacher)	Print Name (Head teacher)
------	--------------------------	---------------------------

Date	Signature (Researcher)	Print Name (Researcher)
------	------------------------	-------------------------

Parents/Carer Information Leaflet

The Nuffield Early Language Intervention

Promoting Oral Language Skills in Nursery/Reception Classes

Researchers based at University College London and the University of Sheffield are carrying out a research project that tests the benefits of an oral language intervention for children in nursery and Reception classes. The programme is called The Nuffield Early Language Intervention and is designed to support children's language abilities so they build a stronger language foundation for school participation.

Your child is being invited to take part in this research project and this leaflet provides you with more details about it.

Please take the time to read the following information carefully and discuss it with others if you wish. It is important for you to understand why the research is being done and what it will involve before you decide whether or not you wish to take part.

The Nuffield Early Language Intervention Programme

The Nuffield Early Language Intervention programme aims to help children in 3 key areas: learning the meanings of words (vocabulary knowledge), speaking skills (using language to convey simple messages/stories) and listening skills (understanding what is said to them). The sessions, are delivered by a Teaching Assistant working in school. The sessions are designed to be fun for children, and all children are encouraged to join in and take an active role. In each session, children will learn new words and take part in a variety of storytelling activities to improve their speaking skills. Finally, children will be rewarded for following a set of "listening rules".



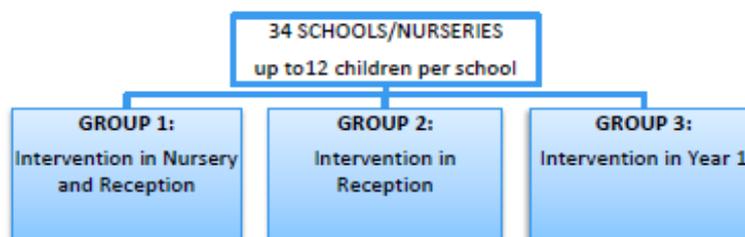
Nuffield Early Language Intervention - Project Update 4

Project Aims:

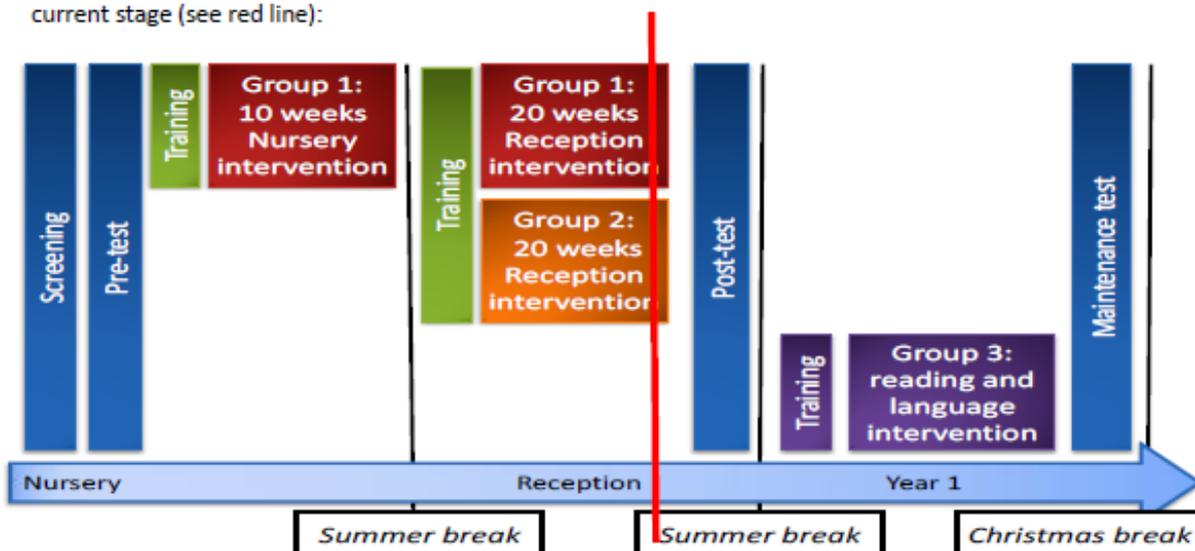
- to compare a 30-week programme of the *Nuffield Early Language Intervention* with a 20-week programme
- to equip school staff with a wide range of skills and material, useful in supporting children with oral language difficulties

Progress to Date:

- Last academic year, members of the project team worked in your school's nursery.
- We selected up to 12 children based on their performance on language screening and pre-test measures to take part in the project. These children were randomly allocated to one of three groups (with up to 4 children per group):



- Following the training by I CAN, the trained teaching assistant (TA) worked with the children in group 1 in nursery and Reception, and also with the children in group 2 in Reception delivering the *Nuffield Early Language Intervention*.
- A member of the research team has visited your school to observe the TA deliver some intervention sessions and to provide feedback.
- We have asked teachers to complete questionnaires about the communication and attention skills of the recruited children (Children's Communication Checklist and the SWAN Scales) and we have asked parents/carers of participating children to complete a short questionnaire to collect some information about the children's language background.
- We have carried out group assessments with children in the year above those participating in the intervention project (i.e. Reception classes before summer break 2013 and Year 1 before Christmas break 2013).
- The figure below gives an overview of the project timeline (from nursery till Year 1 (2012 – 2014)) and the current stage (see red line):





THE UNIVERSITY of York



Next steps until summer break:

- The TA delivering the intervention has been asked to inform the research team about when they will have reached the end of the Nuffield Early Language Intervention programme.

Assessments:

- We will assess all (up to 12) children in the project at two separate time points
 - 1) as soon as possible after the TA finishes delivering the intervention (i.e. post-test)
 - 2) in Oct - Dec 2014 for a follow-up

The post- and follow-up testing phases will involve two members of the research team coming to see these children on an individual basis in a quiet area within your school. The assessments will include age-appropriate language and early literacy tests, and will take place over approximately 2 x 20-30 min sessions per child. In addition, they will be seen as a group. At post-test, this group testing can be combined with the group assessment of children in the year above (see below).

Our project co-ordinators will contact you around the time the TA will finish the intervention delivery to agree on specific dates that are convenient for your setting (likely to be 2 days for the individual assessments).

- We would also like to ask teachers again to complete the questionnaires about the communication and attention skills of the recruited children after the intervention has finished. The research team will bring these with them when they visit the school to assess the children.
- The research team will also repeat the group assessments with children in the year above those participating in the interventions one more time (May – Jun 2014).
- Before summer break, we would like to collect Early Years Foundation Stage (EYFS) Profile scores from the school records for the up to 12 children taking part in the intervention project.

We hope that this 4th project update was informative for you. We are grateful to all schools that are taking part in the study and for their continued support with the project. Please feel free to contact us if you have any questions.

For questions related to the delivery of the Nuffield Early Language Intervention, please contact I CAN:

- Mandy Grist, I CAN Communication Advisor, Tel: 01252 343221, email: mgrist@ican.org.uk

If you have any questions regarding the project's data collection, please contact the research team via one of the project co-ordinators or managers in the north:

- Alexandra Zosimidou, Dept. of Human Communication Sciences, University of Sheffield, 362 Mushroom Lane, Sheffield, S10 2TS. Tel: 0114 222 2410; email: A.Zosimidou@sheffield.ac.uk
- Liam Maxwell, Dept. of Human Communication Sciences, University of Sheffield, 362 Mushroom Lane, Sheffield, S10 2TS. Tel: 0114 222 2458; email: L.Maxwell@sheffield.ac.uk
- Dr Silke Fricke, Dept. of Human Communication Sciences, University of Sheffield, 362 Mushroom Lane, Sheffield, S10 2TS. Tel: 0114 222 2419; email: S.Fricke@sheffield.ac.uk
- Dr Claudine Bowyer-Crane, Dept. of Education, University of York, Heslington, York, YO10 5DD. Tel: 01904 434398; email: claudine.bowyer-crane@york.ac.uk

Appendix E: Topic guide for programme coordinators

Aim of the topic guide for teachers/foundation stage co-ordinators:

Provide a brief overview of how the Nuffield Early Language (NEL) Intervention was set-up and is managed (and delivered in some cases) by teachers in schools; what language provision and support is provided to children in the control group; whether there is any potential for spill-over from the intervention to the control group; and views on the impact of the intervention.

Interviews will be carried face-to-face in 8 schools with teachers or foundation stage co-ordinators. **The interview will last for approximately 30-35 minutes.** The findings from the interviews will contribute to the broader process evaluation of the ICAN language programme/Nuffield Early Language Intervention. It is anticipated that a more detailed exploration of the actual delivery of the programme will be take place in interviews with Teaching Assistants.

This topic guide sets out a number of necessary contextual and factual topics and questions that will be covered during the interview. The guide does not contain follow-up probes and questions like `why`, `when`, and `how`, etc. Researchers will use prompts and probes in order to fully understand how and why views, behaviours and experiences have arisen. The profile information that has already been compiled for each will inform the interview, and guide the questioning and probing.

Introductions (2 minutes)

- **Introduce self & NatCen** (research organisation independent from ICAN & school)
- **Purpose of interview:**
 - This is part of the overall evaluation of the Nuffield Early Language intervention taking place in 32 schools.
 - Study carried out by NatCen in partnership with IFS
 - Commissioned by the Education Endowment Foundation (EEF)
 - In this qualitative part of the evaluation we want to hear about your experiences of delivering the programme. We are interviewing 4 teachers/co-coordinators in other schools
 - This is vital in helping us understand which parts of the programme worked well and which can be improved.
- **Broad topics to be discussed:**
 - *[IF TAS DELIVERING]* We will be discussing the delivery of sessions with TAs. In this interview , we would like to discuss:
 - *Schools experiences of setting up and managing the programme*
 - *The language provision provided to pupils in the control group*
 - *Explore the impacts of the programme on the school and how it fits with other language interventions*
 - *[IF TEACHER DELIVERING INTERVENTION, EXPLAIN WE WILL EXPLORE ALL OF THE ABOVE, AS WELL AS DELIVERY – USE TA GUIDE FOR THIS]*
- **Details of their participation**
 - No right or wrong answers and discussion not a quiz!
 - *Evaluating programme.* We are NOT evaluating the school, only the programme as a whole.
 - *Voluntary.* You don't have to take part and don't have to answer anything you don't want to – free to withdraw from study at any time.
 - *Confidentiality.* Individual & schools not identified in report and we do not share your views with anyone else in school
 - *Obvious questions.* We may ask some very obvious questions, but it is important to hear what you do in your own words.
 - *Length of interview.* Approximately 30-35 mins.
- **Permission to record:** explain you'll be making notes but recording means that we don't have to scribble everything down. **Any questions?**

Phase	Topic & probes/prompts
<p>1. Context and overview of language needs</p> <p>(5 mins)</p> <p><i>Aim: To gather background contextual information about the school, the profile (language and otherwise) of their pupils and the participant's involvement in the programme. [Draw on the profile information to inform discussion]</i></p>	<p style="text-align: center;">A. Role within school and in relation to NEL</p> <p>Overview of their role within the school, including in relation to language support <i>[BRIEF]</i></p> <p>Overview of their role in relation to Nuffield Early Language Programme (NEL):</p> <ul style="list-style-type: none"> • Setting-up the programme • Management/support/facilitating responsibilities <i>[more detail later]</i> • Delivery of the programme – to what extent <hr/> <p style="text-align: center;">B. Understanding and views of NEL</p> <p>Brief description of language interventions in school before NEL?</p> <p>Overall understanding of the aims of NEL <i>[BRIEF]</i></p> <ul style="list-style-type: none"> • What NEL is trying to achieve - esp. rationale & having two interventions <p>General views on the principle of early years language intervention <i>[BRIEF]</i></p> <ul style="list-style-type: none"> • Too soon/too late for children – especially at nursery • Does intervention provide the right structure/too structured an approach to this? • How does what is taught in intervention complement, conflict or replicate • What do they think about the focus of the intervention – Active listening, vocabulary, narrative & phonological awareness
<p>2. Setting up and management of intervention</p> <p>(10 mins)</p>	<p style="text-align: center;">A. School's involvement in NEL</p> <p>Explore history of school's involvement in NEL <i>[PARTICIPANT MAY HAVE LIMITED KNOWLEDGE]</i></p> <ul style="list-style-type: none"> • Reasons for joining programme – including need for it • Previous experience of with similar programmes in previous 3-5 years • Perceptions of need for programme

Aim: To understand how the intervention was set up and is managed in schools; what's involved in the set up in terms of resourcing; and any explore any challenges.

B. Setting up the programme

Experiences and views of setting up the programme.

- Extent of involvement (light / heavy touch; ongoing / one-off tasks)
- Extent of preparation and planning needed to deliver the intervention
 - Timetabling (in place of another class or in addition to)
 - School / classroom space
 - Pressure on school timetable – (e.g. implications of running one to one sessions and number of sessions)
 - Working across nursery and primary schools
- Any training provided/attended?
- Any other early implementation activities?
- Challenges and what worked well
- Any support they received during implementation
 - Sources & type (including ICAN)
 - Views on how helpful/gaps
- Any improvement they would suggest

C. Delivery of intervention

If teacher/coordinator is responsible for delivery as well, merge this guide with the TA guide. If not, only touch on overall views about delivery with participant.

Deciding **who delivers the intervention** (TAs only?)

- How selected/recruited (what considerations, new or existing TA used, what level?)
- Views on the appropriateness of TAs delivering programme (if so, level of TA important? E.g. Advanced TAs)
- Resource implications (e.g. what cover has to be organised)
- Challenges/what worked

D. Ongoing management

Experiences of **on-going management** of the programme

- Extent of involvement (light / heavy touch; ongoing / one-off tasks)
- What on-going management involves
 - Supporting TAs
 - Managing other classroom teachers
- Experiences of programme management – what works well, less well, challenges faced and how overcome
- Any support they received during implementation
 - Sources & type (including ICAN, head teachers, senior staff etc...)
 - Views on how helpful/gaps
- Any improvements they would suggest

E. Resource implications

Explore on-going **resource implications** of the programme

- Staff time (TA, their time, other senior staff time, what cover has to be organised)
- School time (e.g. fitting sessions in school day, week, term)
- Dedicated spaces to deliver group and individual sessions

	<ul style="list-style-type: none"> Managing 2 interventions and control group in their school. Working with nurseries What works/challenging Any improvements they would suggest
<p>3. Support provided to control group & similarities to other support</p> <p>(10 mins)</p> <p><i>Aim: To understand what language support is provided to the control group and the format of the support; and the similarities to other language support provided in the school [Draw on profile information to inform discussion].</i></p>	<p style="text-align: center;">A. Explore what language support pupils in control group receive</p> <p><i>Explain that we understand that there are around 12 children involved in NEL programme in their school – 4 of these are in a 'waitlist control group' and not receiving either the 20 or 30 week intervention.</i></p> <p>Explore what language support these 4 receive in <u>both nursery and reception</u>:</p> <ul style="list-style-type: none"> Type of intervention support/general strategies received <ul style="list-style-type: none"> Nature of support (does the term “integrated reading and language” intervention sound familiar?) If it differs from what the rest of the school receive in nursery and Reception and, if so, how (i.e. business as usual) Similarities/differences to Nuffield language intervention Format and delivery of support [BRIEFLY] <ul style="list-style-type: none"> Who delivers it – is it the same TAs? How it is delivered (small group, one-to-one etc...) Frequency and timing How delivered What support do children in the 20 week intervention received whilst in nursery <p style="text-align: center;">B. Other language interventions taking place in school</p> <p>Explore other language interventions taking place at nursery and Reception [BRIEFLY]</p> <ul style="list-style-type: none"> What these are How do they sit with NEL? (E.g. reinforce it, replicate it, are separate from it) <p style="text-align: center;">C. Potential for spill-over</p> <p>Views on the potential spill-over of ICAN teaching methods to the control group (and other pupils in the school)</p> <ul style="list-style-type: none"> Explore whether TAs delivering NEL are a dedicated resource or deliver other classes If deliver other classes, explore extent to which they feel NEL teaching approach is being used in these other classes <ul style="list-style-type: none"> If not, why not? If so, how so? (E.g. teaching techniques, materials, lesson structures) etc... The impact this has on these other classes (e.g. raises language levels etc...)
<p>4. Impacts</p> <p>(5 mins)</p> <p><i>Aim: To explore the impact, positive and otherwise, that the programme has had and why</i></p>	<p style="text-align: center;">A. Impacts on Pupils</p> <p>Explore impacts on pupils</p> <ul style="list-style-type: none"> The <i>types of impact</i> <ul style="list-style-type: none"> Language & literacy Wider impacts (e.g. soft skills, other classes) Overall school performances (e.g. missing out on classes) The <i>sustainability of impact</i> – long-term or short-term <i>If none</i>, why not? <i>Who</i> has the programme impacted most and least? Reasons for this. Explore <i>which aspects of the programme</i> critical for each of these impacts <ul style="list-style-type: none"> Structure (e.g. TA interaction, duration, how long lessons last etc...) Sessions and elements within sessions Teaching principles (e.g. scaffolding, progressing etc...) How programme is delivered (e.g. child being taken out of class, same TA, different TA) Explore other <i>non-programme related factors</i> that contributed to observed impacts (e.g. parents)

	<ul style="list-style-type: none"> • Explore which NEL intervention(s) had the most impact and why (e.g. control, all 30 weeks, Reception 1, Reception 2) • How impact can be <i>improved</i> <ul style="list-style-type: none"> ○ Probe around children dropping out (especially those moving from nursery to a reception class in another school). <p style="text-align: center;">B. Impacts on staff and school</p> <p>Impacts on staff and school</p> <ul style="list-style-type: none"> • If not, why not? • Types of impact <ul style="list-style-type: none"> ○ On staff - both those directly involved in NEL (TA's / teachers) and others. E.g. <ul style="list-style-type: none"> ▪ Job satisfaction ▪ Workloads ▪ Staff resources ○ Impact on school language programme ○ Impact on relationship between nursery and primary school • Why programme has had impact – which aspects of programme have contributed to this <p style="text-align: center;">C. Other impacts</p> <p>Any other impacts?</p> <p>Parents</p> <ul style="list-style-type: none"> • Parent-child relationships • Parent-school relationships • Their own literacy <p>Others</p> <p>For each, probe on</p> <ul style="list-style-type: none"> • Nature of impact • Which aspects of programme contributed to these <p>How impact can be improved</p>
<p>5. Views and reflections</p> <p>(3 mins)</p> <p><i>Aim: To explore views about rolling out the programme with all pupils; and any concerns over value for money</i></p>	<p style="text-align: center;">A. Value for Money</p> <p><i>Note that that they may not be aware of the cost of the programme and cost of inputs for the programme.</i></p> <p>Views on the value for money of the interventions</p> <ul style="list-style-type: none"> • Do they think it is a cost-effective way of delivering language intervention <p>If they had to pay for the programme, would they still consider using it in their school?</p> <ul style="list-style-type: none"> • If not, why not? • If so, <ul style="list-style-type: none"> ○ Which intervention(s) would they use and why? (e.g. 30 week, 20 week or even the control)

	<ul style="list-style-type: none"> ○ How would they fund the intervention (<i>given that it will no longer be free</i>)
<p>6. End</p>	<p style="text-align: center;">B. Final Reflections</p> <p>Final reflections on the programme</p> <ul style="list-style-type: none"> • Strengths of programme • Challenges • Overall suggestions for change <ul style="list-style-type: none"> • Ask if anything else to add • Reassure about confidentiality • Stop recording • Thank them for their time

Appendix F: Topic guide for teaching assistants

Aims of teaching assistant topic guide

- **Preparedness.** An understanding of the training and support received by teaching assistants and what can help/did help them feel prepared for delivery.
- **Understanding actual delivery within schools.** A rich description of how the programme is delivered in schools, including:
 - The structure of the programme
 - Delivery of sessions – including elements within sessions and teaching techniques used
 - Description of delivery within the school context
 - Fidelity to the programme – i.e. how delivery reflects
- **Evaluating the delivery of the programme.** Views on the delivery that can help inform any wider roll-out of the programme. This includes understanding factors that:
 - Facilitated/hindered delivery
 - What can improvements be made to the delivery
- **Reflecting on the impact of the programme.** Views on the impact of the programme will help clarify and contextualise the quantitative findings. Focus will be:
 - The types of impacts programme is perceived to have had– if any – including reasons
 - Who the programme has had the most/least impact on
 - Which aspects of the programme are critical to the impact (which interventions, elements etc...)
 - What other, non-programme factors have contributed to the impact.

By the end of the interview, we should have a clear idea of how the programme runs, respondents' thoughts on its implementation and impact. Extensive fieldnotes will be written up afterwards.

Introduce self & NatCen (as a research organisation independent from I CAN & school)

Purpose of interview:

- This is part of the overall evaluation of the Nuffield Early Language intervention taking place in 33 schools.
- Commissioned by the Education Endowment Foundation (EEF). Study carried out by NatCen in partnership with IFS
- In this qualitative part of the evaluation we want to hear about your experiences of delivering the programme. We are interviewing 8-16 TAs in other schools also.
- This is vital in helping us understand which parts of the programme worked well and which can be improved.

Broad topics to be discussed: We will discuss [Nursery and/or Reception interventions]. Broad areas of discussions are:

- Training and support you received
- How programme is delivered in your school and their views on delivery
- Any reflections you have on how effective the programme is

Details of their participation

- No right or wrong answers and discussion not a quiz!
- *Evaluating programme.* We are NOT evaluating the school, only the programme as a whole.
- *Voluntary.* You don't have to take part and don't have to answer anything they don't want to – free to withdraw from study at any time.
- *Confidentiality.* Individual & schools not identified in report and we do not share your views with anyone else in school
- *Obvious questions.* We may ask some very obvious questions, but it is important to hear what you do in your own words.
- *Length of interview.* 1 hour. **Permission to record:** explain you'll be making notes but recording means that we don't have to scribble everything down. **Any questions?**

Phase	Topic & probes/prompts
-------	------------------------

<p>1. Context</p> <p>(5 mins)</p>	<p style="text-align: center;">A. Their involvement in the programme</p> <p>Explore their current role in the school [briefly]</p> <ul style="list-style-type: none"> • How long been in school • Their current role and responsibilities • Any history of delivering language teaching <p>Their specific role in relation to the Nuffield Early Language Intervention (NEL) - [briefly – cross-check with our profiling information] – <u>including if they work with children in the control/in-waiting group?</u></p> <p>Their overall understanding of what the NEL is trying to achieve [briefly] – esp. rationale & having 2 interventions</p> <hr/> <p style="text-align: center;">B. Interventions prior to NEL</p> <p>Describe (if any) language interventions that took place prior to NEL at Nursery and Reception level [briefly] –</p> <ul style="list-style-type: none"> • If none, what gap (if any) NEL fills • Brief description of aims of prior intervention(s) • How these compares to NEL (better/worse; NEL different or replicates)
<p>2. Training & ongoing support</p> <p>(10 mins)</p>	<p style="text-align: center;">A. Initial training</p> <p>Explore experiences of training to deliver the Nursery intervention [ONLY if they delivered it in Nursery]</p> <p>This should be a 1 day training that took place around March 2013; but check what they received.</p> <p>Explore experiences of training to deliver the Reception intervention [ONLY if they delivered it in Nursery]</p> <p>This should be a 2 day training that took place around September 2013; but check what they received.</p> <p>For each, probe around:</p> <ul style="list-style-type: none"> • Description of training (when, content and workload) [Briefly] • How prepared they felt to deliver NEL at end of training and why <ul style="list-style-type: none"> ○ Most useful aspects of training & reasons – including if training duplicates what they have done on the past ○ Gaps in training ○ Impact of training/gaps on delivery ○ How training could be improved <hr/> <p style="text-align: center;">B. Ongoing training</p> <p>Experiences and views on any formal on-going support they received from I CAN.</p> <ul style="list-style-type: none"> • Description of support. Check whether they received any of the following and, if so, a) for how long and b) the type of support these offered <ul style="list-style-type: none"> ○ Telephone support (30 weeks) ○ Online forums ○ I CAN manual • Impact of support/lack of support on delivery • Gaps in support and how they would improve it

	<p>Experience of other forms of support they may have received outside of ICAN</p> <ul style="list-style-type: none"> • Source (e.g. school, teachers, other TA's), description of support (formal, informal etc...) & amount of support • Impact of support/lack of support on delivery • Gaps and improvements
<p>3. Delivery – mapping and views (40 mins)</p>	<p style="text-align: center;">A. Description of overall delivery structure</p> <p>This section is about getting an overview of how the programme is delivered in the 10 weeks in nursery and 20 weeks in Reception. Explain we will talk about the delivery of individual sessions shortly.</p> <p>Explore briefly the overall delivery structure of the intervention at nursery. [ONLY if they delivered it in Nursery]</p> <p>Explore briefly the overall delivery structure of the intervention at Reception. [ONLY if they delivered it in R]</p> <p>Allow participants to describe spontaneously and then probe on the following, looking at the rationale also.</p> <ul style="list-style-type: none"> • <i>Duration of intervention</i> in school term [should be 10 weeks in nursery & 20 in R, covering 30 sessions] • <i>How often</i> a week sessions are delivered and on which days – rationale for this [at least 3 group sessions delivered a week, although days can vary. At R, alternate bw group and individual sessions] • <i>Who</i> delivers it [should be TA] • <i>Where</i> delivery takes place • <i>When</i> intervention is delivered (e.g. during certain lessons, is it the same lessons or different ones?) • <i>Number of children</i> seen during sessions [maximum of 4] • <i>When key milestones</i> are reached <ul style="list-style-type: none"> ○ Initial assessments take place [Nursery - should be first week and week 10; R1 – 2 assessments week 1 and 2 in week 10; R2 – 1 assessment week 1 and 2 in week 10] ○ When consolidation takes places [Nursery - should be week 30; R1 – session 5,10,14,19,24,28; R2 – session 10,20,29] • The <i>topic areas</i> covered [Nursery - family & friends and our house; R1 – My body, things we wear, people who help us; R2 – growing, journey & time] <p style="text-align: center;">B. Explore how typical group sessions are delivered</p> <p>Explain we are going to move on to talk about the delivery of sessions</p> <p>Ask them to briefly explain how typical group sessions are delivered at nursery [ONLY if they delivered it in Nursery]. In addition to typical delivery, explore what happens during the following key sessions:</p> <ul style="list-style-type: none"> • <i>Consolidation</i> session in week 10 <ul style="list-style-type: none"> ○ How are parents informed ○ What are they told ○ Are they encouraged to do activities with child? • <i>Assessment</i> sessions <ul style="list-style-type: none"> ○ Is the Nursery Progress Assessment (Taught Vocabulary) tool used both at the start and end? Do they also repeat the assessment at other times to suit their needs? ○ Can they find a quiet space to do the assessment? ○ Do they use the taught vocabulary record sheet? ○ Who test is administered and scored by? <p>Explore briefly whether typical group sessions delivered at Reception differ from nursery [ONLY if they delivered it in Reception]. As above, also explore what happens during consolidation and assessment sessions. For assessments, if and when a) Reception Programme Progress Assessment</p>

(Taught vocabulary) used and b) for R2, is Progress Assessment (Taught Letter Sounds) tool used.

Allow participants to describe spontaneously and then probe on the following:

- *Preparation* – time taken [should be 1 hr in Nursery and 2 hours in R] and what it involves
- *Duration* of sessions [groups should be 20 mins at Nursery & 30 mins at R]
- *The key elements of the sessions* – a) what these are b) the order in which these appear c) what is covered d) how long do they last
- Teaching techniques and approaches used [recasting/corrective feedback; modelling; highlighting; prompting; scaffolding; multisensory learning; progressing; consolidating]
- Key decisions made *within elements*. E.g. at nursery
 - Intro – any icebreaker sessions they use (e.g. 'hide the teddy')
 - Listening game - did they draw on the Listening Games Bank or do they devise or modify these?
- How pupil progress monitored.
 - Nursery – whether on-going records kept during the 10 week
 - R1 & R2 – should use **Planning & Record Sheets**

C. Explore how 1-2-1 sessions are delivered at R1 & R2

Explore **briefly** how **individual session** delivered in Reception 1 and 2. In addition to overall delivery, explore how sessions are **tailored** to child's needs [these are designed to focus on child's individual needs].

- How tailored (confident/not confident words and narratives)
- What used to decide how sessions should be tailored? (ideally, notes and observations)
- Does the structure of the session enable it to be tailored to meet child's need?

For both, probe on:

- *Preparation time* [2 hours in total a week for both R1 & R2 & use of planning sheets]
- *Duration* [1-2-1 = 15 mins]
- *Key elements* – a) what these are; b) the order in which they appear c) What is covered; d) how long they last
- *How pupil progress is monitored* in an on-going way [use of should use planning and record sheets]
- *Teaching techniques and approaches* used during the sessions [same as groups – e.g. scaffolding etc...]

D. Fidelity [IMPORTANT FOR IMPACT EVALUATION]

How far does the delivery of NEL **mirror their training/manual**? Explore deviations at Nursery, R1 & R2 (groups and 1-2-1 sessions) in:

- *Overall structure* (e.g. number of weeks it lasts, number of sessions delivered, whether there is the danger of 20 and 30 week have intervention delivered together in Nursery)
- *Within sessions* (e.g. in how each element is delivered, the order in which they are delivered, the duration of sessions, length of each element etc...)
- *Teaching approach*
- *Rationale for any deviations* (needs of individual child, resource issues etc...)
- *Barriers/facilitators* to fidelity
- *If no deviations, why not?*

E. Evaluating delivery – both structure and sessions

Overall views on delivering the **nursery intervention** – [ONLY if they delivered it in Nursery] including views on both overall structure and delivery of sessions.

- *Any changes* in experience of delivering sessions over time
- What *works well* and why
- Nature of any *challenges* making delivery work in their school setting

	<ul style="list-style-type: none"> • Whether/how challenges <i>overcome</i> • <i>How intervention meets its teaching goals (aims to be enjoyable, child friendly & gives TAs skills)</i> • Any suggested <i>improvements for roll-out</i> <p>Views on transition of programme from Nursery to Reception - [BOTH Nursery & Reception TAs will have a view on this] including views on both overall structure and delivery of sessions.</p> <p>If different TA delivering it at nursery, views on the process of transition?</p> <ul style="list-style-type: none"> • What the process of transition involved (<i>briefly</i>) • What worked • What was a challenge and how overcome <ul style="list-style-type: none"> ◦ Probe around dropouts between Nursery & Reception, an issue in London esp. • What can be improved/done differently <p>If same TA, views on this</p> <ul style="list-style-type: none"> • How the transition to Reception worked • What were advantages of this • What were the disadvantages • What can be improved/done differently <p>Overall views on delivering group sessions Reception in R1 & R2– including views on both overall structure and delivery of sessions.</p> <ul style="list-style-type: none"> • <i>Any changes</i> in experience of delivering sessions over time (e.g. <i>how views about introducing the phonics element in R2</i>) • What <i>works well</i> and why • Nature of any <i>challenges</i> making delivery work in their school setting • Whether/how challenges <i>overcome</i> • <i>How intervention meets its teaching goals (aims to be enjoyable, child friendly & gives TAs skills)</i> • Any suggested <i>improvements for roll-out</i> <p>Overall views on delivering the 1-2-1 sessions in R1 & R2 – including views on both overall structure and delivery of sessions.</p> <ul style="list-style-type: none"> • <i>Any changes</i> in experience of delivering sessions over time • What <i>works well</i> and why (<i>does it work to alternate group & 1-2-1 sessions, do sessions help to deepen knowledge/help children to catch-up?</i>) • Nature of any <i>challenges</i> making delivery work in school setting • Whether/how challenges <i>overcome</i> • <i>How intervention meets its teaching goals (aims to be enjoyable, child friendly & gives TA skills)</i> • Any suggested <i>improvements for roll-out (Should they have 1-2-1 sessions in nursery too?)</i> <p>FOR ALL - Probe around factors that affect delivery:</p> <ul style="list-style-type: none"> • NEL related (NEL teaching, the tools/equipment [e.g. listening games, flash cards, Ted], pace of teaching, no. of sessions to be delivered, meeting administrative duties, challenges specific to each element – e.g. listening, vocabulary, narrative, phonics, plenary) • School related factors (e.g. other staff, school capacity & resources, timetable) • Staffing issues (No. of TAs delivering, appropriateness of TA delivering it) • Time related factors (fitting intervention into 20 minutes, preparation time, fitting 30 sessions in 10 weeks) • Pupil related factors (appropriateness of pupils to intervention, attendance, engagement) • Parent related factors (level of engagement with child's education)
<p>4. Impact (10 mins)</p>	<p style="text-align: center;">A. Impact on pupils</p> <p>Explore any (positive or otherwise) impact the programme had on pupils and why</p>

	<ul style="list-style-type: none"> • The <i>types of impact</i> <ul style="list-style-type: none"> ○ Language & literacy ○ Wider impacts (e.g. soft skills, other classes) ○ Overall school performances (e.g. missing out on classes) • The <i>sustainability of impact</i> – long-term or short-term • <i>If none, why not?</i> • <i>Who</i> has the programme impacted most and least? Reasons for this. • Explore <i>which aspects of the programme</i> critical for each of these impacts <ul style="list-style-type: none"> ○ Structure (e.g. TA interaction, duration, how long lessons last etc...) ○ Sessions and elements within sessions ○ Teaching principles (e.g. scaffolding, progressing etc...) ○ How programme is delivered (e.g. child being taken out of class, same TA, different TA) • Explore other <i>non-programme related factors</i> that contributed to observed impacts (e.g. parents) • Explore which NEL intervention(s) had the most impact and why (e.g. control, all 30 weeks, Reception 1, Reception 2) <ul style="list-style-type: none"> ○ Did Reception TA notice difference bw 20 & 30 week group at start of Reception? ○ Did these differences change over time – e.g. 30 week better prepared at start of Reception bt differences did not last/continued? • How impact can be <i>improved</i>
	<p>B. Other impacts</p>
	<p>School and school staff. Probe particularly on:</p> <ul style="list-style-type: none"> • If not, why not? • Types of impact <ul style="list-style-type: none"> ○ On staff - both those directly involved – esp. TA's themselves. E.g. <ul style="list-style-type: none"> ▪ Job satisfaction ▪ Workloads ▪ Staff resources ○ Impact on school language programme ○ Impact on relationship between nursery and primary school <p>•Why programme has had impact – which aspects of programme have contributed to this</p> <p>Parents</p> <ul style="list-style-type: none"> • Parent-child relationships • Parent-school relationships • Their own literacy <p>Others</p> <p>For each, probe on</p> <ul style="list-style-type: none"> • Nature of impact • Which aspects of programme contributed to these • How impact can be improved
<p>5. End</p>	<ul style="list-style-type: none"> • Ask if anything else to add • Reassure about confidentiality • Stop recording • Thank them for their time

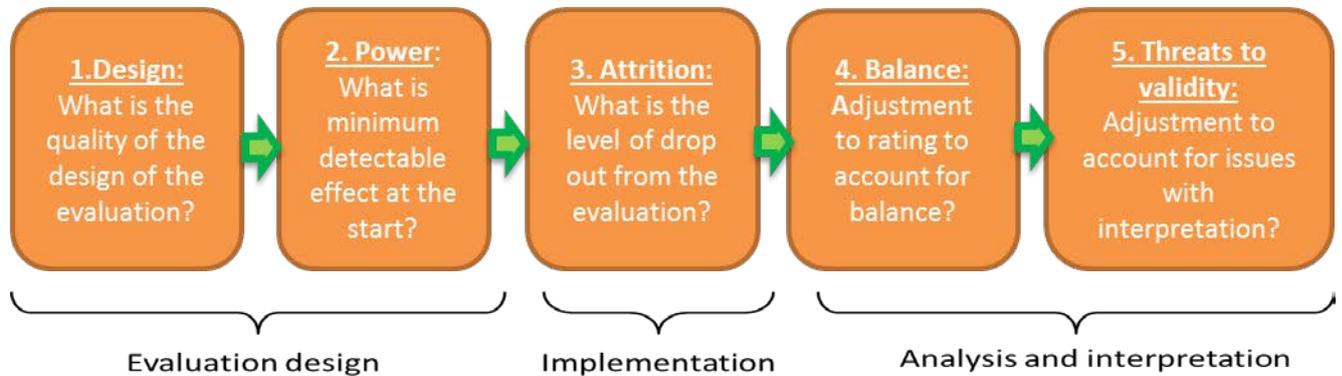
Appendix G: Cost rating

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

Appendix H: Security classification of trial findings

25 February 16 Completed by Camilla Nevill



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

The final security rating for this trial is 4 . This means that the conclusions have moderate to high security.

The trial was designed as an efficacy trial and could achieve a maximum of 4 . There was low attrition among pupils of 11% and only small differences in prior attainment between arms. There is some threat to validity of the findings because of the risk of spillover within the schools to the control TAs. However, there is evidence from the process evaluation that this is not a substantial threat.

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

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

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

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



Education
Endowment
Foundation

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